LNCS 3160

Stephen Brewster
Mark Dunlop (Eds.)

Mobile Human-C Interacti MobileHC

6th International Symposi
Glasgow, UK, September 20
Proceedings

Stephen Brewster Mark Dunlop (Eds.)

Mobile Human-Computer Interaction – MobileHCI 2004

6th International Symposium, MobileHCI 2004
Glasgow, UK, September 13 - 16, 2004
Proceedings

Springer

Volume Editors

Stephen Brewster
University of Glasgow, Department of Computing Science
Glasgow, G12 8QQ, UK
E-mail: stephen@dcs.gla.ac.uk

Mark Dunlop
University of Strathclyde, Department of Computer and Information Sciences
Livingstone Tower, Richmond Street, Glasgow G1 1XQ, UK
E-mail: Mark.Dunlop@cis.strath.ac.uk

www.mobilehci.org

Library of Congress Control Number: 2004111167

CR Subject Classification (1998): H.5.2, H.5.3, H.5, H.4, C.2, I.2.1

ISSN 0302-9743
ISBN 3-540-23086-6 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Protago-TeX-Production GmbH
Printed on acid-free paper SPIN: 11321507 06/3142 5 4 3 2 1 0

Preface

MobileHCI is a forum for academics and practitioners to discuss the challenges and potential solutions for effective human-computer interaction with mobile systems and services. It covers the design, evaluation and application of techniques and approaches for all mobile computing devices and services. MobileHCI 2004 was the sixth in the series of conferences that was started at Glasgow University in 1998 by Chris Johnson. We previously chaired the conference in 1999 in Edinburgh (as part of INTERACT 1999) and in 2001 in Lille (as part of IHM-HCI 2001). The last two years saw the conference move to Italy, first under the chairmanship of Fabio Paternò in Pisa then under Luca Chittaro in Udine. In 2005 the conference will move to Austria to be chaired by Manfred Tscheligi. Each year the conference has its own website hosted by the conference chair, however the address www.mobilehci.org will always point to the next (or current) conference.

The number of submissions has increased every year. This year we received 79 full papers (63 were received last year) from which we accepted the best 25. We had 81 short papers and posters submitted (59 last year) and accepted 20 of these as short papers and 22 as posters. We received 9 workshop, 4 tutorial and 2 panel proposals, from which 5, 2 and 2, respectively, were accepted.

All papers were reviewed by two reviewers, and any papers where the reviewers' ratings were widely different were reviewed a third time. This allowed us to keep the quality of the work presented very high. We would like to thank all of the reviewers for their help and time. It was great to see so many people put so much of their time into the conference. The quality of the reviews helps the field as a whole get better and better.

Traditionally there has been a split at MobileHCI, with much of the academic research presented being carried out on palmtop devices while industrial interest has always been stronger from mobile telephone companies. While this is still true for MobileHCI 2004, the state-of-practice is showing strong convergence with many phones now being powerful handheld computers too. The presentations at MobileHCI 2004 covered a broad range of research into the usability of mobile devices in the widest sense, reflecting the very different nature of interaction on mobile devices compared to traditional desktop interfaces. This year we identified four main themes within the papers: overcoming device limitations; evaluating mobile systems; supporting diverse user groups; and the mobile Web.

The first theme looks at two fundamental problems with small devices: input and output. Interaction techniques are restricted because we want a small device that fits comfortably into a pocket; furthermore power is a serious problem with battery technology often lagging behind the power needs of researchers and limiting or influencing how devices work in practice. Papers in these sessions looked at novel interface designs to handle these problems with small devices, including increased use of sound, use of tilt and gesture, detailed investigations

of very small display areas for meaningful interaction and the design of interfaces to better support users of power-limited devices.

Our second theme moves into another key area of difference between traditional office/desktop-based HCI and mobile HCI: evaluation. Trials of users sitting at a desk doing a fixed set of tasks without interruption just do not feel appropriate for assessing interfaces that will be used while users are walking around, at home or otherwise enjoying themselves. The papers in these sessions looked at whether this intuition about mobile evaluation is correct and at evaluations conducted in different settings.

Access to the World-Wide Web has become a core part of many people's lives at home and in the office. Providing access on small devices is very challenging as the pages are almost universally designed for large desktop screens. A number of papers discuss different approaches to providing Internet access on mobiles, including transforming the pages visually, personalization of pages and making more use of different modes of interaction.

Office workers, the core of much of traditional HCI, are a fairly homogeneous group of people, but mobile devices are being used by a much wider population. A small group of papers at MobileHCI 2004 looked at different user groups, particularly older and younger users. There are many potential users in these categories but little research has been done on their needs, wants and capabilities.

The proceedings are split into ten sections, the first seven covering the full papers with the remainder covering the other submission categories of short papers, posters, tutorials and workshops, and panels.

Finally, we extend our gratitude to the many people who worked to make MobileHCI 2004 happen and to our sponsors for their generous support of the conference. Thanks also go to the student volunteers who did a great job making things run smoothly.

September 2004

Stephen Brewster and Mark Dunlop
Chairs
MobileHCI 2004
www.mobilehci.org

Organization

MobileHCI 2004 was organized by the Department of Computer and Information Sciences, University of Strathclyde, UK and the Department of Computing Science, University of Glasgow, UK.

Conference Committee

Conference Chair	Mark Dunlop (University of Strathclyde, UK)
Programme Chair	Stephen Brewster (University of Glasgow, UK)
Treasurer	Steven Neely (University of Strathclyde, UK)
Tutorials and Workshops Chair	Marilyn McGee (University of Glasgow, UK)
Panels Chair	Matt Jones (University of Waikato, New Zealand)
Student Volunteers Chair	Joy Goodman (University of Glasgow, UK)
Web Support	Gary Gray (University of Glasgow, UK)
Student Volunteers	Georgios Marentakis (University of Glasgow), Paul Tennant (University of Glasgow), John Williamson (University of Glasgow), Nick Bradley (Strathclyde University), Andreas Komninos (Strathclyde University), Simon Sweeney (Strathclyde University), Parisa Eslambolchilar (Hamilton Institute), Steven Strachan (Hamilton Institute)

Referees

Russell Beale, University of Birmingham, UK
Robert Biddle, Carleton University, Canada
Barry Brown, University of Glasgow, UK
Lorna Brown, University of Glasgow, UK
Matthew Chalmers, University of Glasgow, UK
Alan Chamberlain, Loughborough University, UK
Keith Cheverst, Lancaster University, UK
Luca Chittaro, University of Udine, Italy
Andy Cockburn, University of Canterbury, New Zealand
Joelle Coutaz, CLIPS IMAG, France
Murray Crease, NRC, Canada
Alan Dix, Lancaster University, UK
Philip Gray, University of Glasgow, UK
Tom Gross, Bauhaus-University Weimar, Germany

Ken Hinckley, Microsoft, USA
Simon Holland, Open University, UK
Christopher Johnson, University of Glasgow, UK
Matt Jones, University of Waikato, New Zealand
Steve Jones, University of Waikato, New Zealand
Anne Kaikkonen, Nokia, Finland
Pekka Ketola, Nokia, Finland
Panu Korhonen, Nokia, Finland
Vuokko Lantz, Nokia, Finland
Peter Lonsdale, University of Birmingham, UK
Joanna Lumsden, NRC, Canada
Gary Marsden, University of Cape Town, South Africa
Masood Masoodian, University of Waikato, New Zealand
David McGookin, University of Glasgow, UK
David Morse, Open University, UK
Roderick Murray-Smith, University of Glasgow, UK
Laurence Nigay, CLIPS IMAG, France
Bruno von Niman, von Niman Consulting, Sweden
Eamonn O'Neil, University of Bath, UK
Reinhard Oppermann, Fraunhofer FIT, Germany
Fabio Paternò, ISTI-CNR, Italy
Antti Pirhonen, University of Jyväskylä, Finland
Mattias Rauterberg, Technical University Eindhoven, Netherlands
Thomas Rist, DFKI, Germany
Boris de Ruyter, Philips, Netherlands
Daniel Salber, IBM T.J. Watson Lab, USA
Chris Schmandt, MIT MediaLab, USA
Albrecht Schmidt, University of Munich, Germany
Matthias Schneider-Hufschmidt, Siemens, Germany
Andrew Sears, University of Maryland, USA
Manfred Tscheligi, CURE, Austria
Paul Vickers, University of Northumbria, UK
David Williams, Motorola, Italy

Sponsors

Platinum Sponsor

www.nokia.com

Gold Sponsor

www.microsoft.com

Bronze Sponsor

www.kelvininstitute.com

Other support was provided by the EPSRC-funded AudioClouds project (www.audioclouds.org). We are grateful for the support of Strathclyde and Glasgow Universities and the Glasgow Interactive Systems Group.

Table of Contents

Full Papers: Tilt, Touch, and Text Entry

Full Papers: Auditory Interactions

Full Papers: Device Differences and Web Pages

Full Papers: Novel Interaction Techniques

Short Papers

Posters

Tutorials and Workshops

Panels

What Can You Say with Only Three Pixels?

Christopher Campbell[1] and Peter Tarasewich[2]

[1] IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120 USA
ccampbel@almaden.ibm.com
[2] College of Computer and Info. Science, Northeastern University, Boston, MA 02115 USA
tarase@ccs.neu.edu

Abstract. The size limitations of mobile devices can make information display especially difficult. Micro-displays must take into account the viability of different sizes and configurations for informing users, the flexibility they provide for different types of messages, and under which conditions these results are achieved. An experiment was performed to measure user learning and comprehension of five sets of messages of increasing information size and complexity on a simulated three-light visual display. Results show that these "pixel-based" displays can transmit detailed, information-rich messages up to 6.75 bits in size with minimal training.

1 Introduction

Computer devices and systems are no longer constrained to relatively permanent work or home environments, but are likely to be found in almost any physical location or in any social setting. In such technology-laden environments, information overload can be a serious problem. While information is necessary to perform many tasks, the human mind is limited in terms of how much information it can process at one time. The problem of information management becomes even more difficult and complex in mobile environments. One way to reduce information overload is through the use of meta-information, which can require less effort to process and can result in fewer or less severe disruptions. If meta-information is deemed important, the person receiving it can make a decision whether or not to seek additional details. For example, a mobile worker may not need or want the entire contents of a message every time one becomes available. It may be too distracting (or too dangerous) to the worker's primary tasks. However, they may wish to receive a notification that a message is available, along with an indication of how important it is, and its source. That way, the worker can make their own decision, based on their current situation, whether or not to stop their primary task to access the contents of the message.

To be successful, notification systems (and cues) must "present potentially disruptive information in an efficient and effective manner to enable appropriate reaction and comprehension" [10]. Notification cues are a form of information, and questions arise such as what form these cues should take and how appropriate they are in different settings. Determining notification cues for use in ubiquitous environments can become quite complex, requiring the selection of appropriate delivery channels based on continuously changing contexts and dynamic information needs [9].

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 1–12, 2004.

This paper presents results from an experiment that measured the learning and comprehension of visual notification cues that conveyed increasing amounts of information to the user. Three lights were used, each with three colors and two intensity levels. A previous study [14] showed that the three-light design (compared to other designs with more or fewer lights) was a good choice for conveying notifications on small devices. But that study tested only a fixed amount of information. This experiment extends that work and investigates whether or not larger amounts of information can be conveyed using the same three-light design, and how well people can learn to use the notification cue itself.

2 Background

Notification cues can be visual, auditory, tactile, or multimodal in nature. They can be private such that only the receiver is aware of them, or public such that everyone in the immediate vicinity will receive the cue. Cues can also range from being quiet and subtle to being loud and intrusive. A ringing cell phone is an auditory, intrusive, and very public notification cue. A vibrating cell phone is a tactile, subtle, and private notification cue that can convey the same information [1].

The design and use of notification cues must take into account the intricacies of human attention, which involves the allocation of perceptual or cognitive resources to something at the expense of not allocating them to something else [2]. Humans have a limited amount of resources available for allocation to different tasks and cannot attend to everything at once. People can attend to a modality (e.g., vision, hearing, touch), a color, a shape, or a location [2]. The decision to attend specifically to one of these over the others arises from the task at hand.

With computer applications that are used in the office, home, or similar settings, the context is known and is relatively stable from minute to minute. While this does not mean that there cannot be multiple activities competing for a user's attention (e.g., animated ads and email notifications), the user's environment outside of the computer is fairly consistent from day to day. Most offices and homes function with a fair amount of regularity and predictability, even if they experience a lot of activity. The user can devote relatively consistent attention to performing tasks on the computer.

On the other hand, with mobile applications, there can be a significant number of people, objects, and activities vying for a user's attention aside from the application or computer itself [13]. Furthermore, since devices are completely mobile, this outside environment can change rapidly. A mobile application may not be the focal point of the user's current activities [3], as the user may be trying to balance interaction with a mobile device with other elements in the environment (e.g., walking along a busy city street with small children while receiving directions from a navigation system). Mobile activities can be complex because of changing interactions between the user and the environment. The amount of attention a user can give to a mobile application will vary over time, and a user's priorities can also change unpredictably [5].

An environment that consists of too many distractions can be confusing and unmanageable. Notification cues must be designed such that they minimize the possibility of overloading the attention of the intended recipient and any surrounding people. Otherwise, the cues may prove to be ineffective or ignored completely.

Much effort has been devoted recently to studying notification systems in the form of secondary displays, or peripheral displays, which provide information to the user that is not central or critical to their current or primary task. For example, current news headlines may scroll across a one-line display on a computer screen. Studies have looked at the effectiveness of presenting information in secondary displays in various formats [8, 10]. Research has generally found that performance on primary tasks is negatively impacted by secondary tasks [8], with some exceptions [10].

Other research has investigated notification systems and devices specifically for mobile environments. Wisneski [15] described a subtle and private notification device in the form of a watch that changes temperature as stock prices change. Holmquist, Falk, and Wigström [4] tested a device called the "hummingbird" that notified its user of the close proximity of other group members by producing a sound ("humming") and listing identities of the group members.

In a first attempt at creating a subtle notification cue that was also public, Hansson and Ljungstrand [1] created a "reminder bracelet", worn on the user's wrist, which notifies a user of upcoming events (e.g., meetings). The bracelet consists of three red LEDs that are triggered progressively as an event draws closer. With this device, people near the user can clearly see that the user is being notified about something.

Tarasewich et al. [14] conducted a study that measured the performance/size tradeoff of visual displays that ranged in size from two lights to nine lights, and used display characteristics, such as color and blinking, in various combinations. Results showed a reliable tradeoff between performance (response time and accuracy) and display size (number of lights). However, even the full set of twenty-seven messages used in the study could be conveyed with high recognition accuracy using only three lights by mapping the messages into color and position. The authors concluded that mobile devices with micro-level form factors could be designed to convey critical information and provide effective notifications. However, two issues were not explored in this study. One issue was how learning affected the comprehension and use of the visual displays. The second was how much information could be effectively conveyed using a display of a given size.

3 Evaluating Increasing Information Amounts

Mobile notification display designs should quickly and completely inform users on a small form factor without requiring a lot of attention or training. While small screens exist (e.g., watches), lower information rate displays such as LEDs have the benefit of (a) requiring less cognitive effort to understand (i.e., less distraction), (b) allowing for smaller and even micro level form factors (e.g., jewelry), and (c) using less power. Simply speaking, the less information conveyed, the less attention required to use it. However, less information does not mean the message is not informative. Even small amounts of critical information can be highly informative in mobile situations.

An experiment was conducted to test comprehension and learning of visual cues conveying increasing amounts of information on a three-light display. [14] showed that a three-light design has a balance of good user performance and high user preference, all within a relatively small footprint. The mappings were chosen to be as simple and direct as possible using color and/or position for the information categories and values. The goals of the present experiment were to determine (1) how

well users can progressively learn increasingly complex messages on a three-light display, and (2) how much information can be conveyed successfully on that display.

Table 1. Cue categories and associated values

Category	Possible Values
Source	family, friends, work
Medium	email, voicemail
Type	new, reply, forwarded
Length	long, short
Priority level	high, medium, low

3.1 Information Mapping Functions

The same physical display size was used for each round of the experiment. Each simulated light could show the colors red, blue, and green. Each color could also display at one of two intensity levels – low or high (i.e., dim or bright). This means that we could theoretically encode six pieces of information on a single light (3 colors x 2 intensity levels). With three lights, we can encode a maximum of 216 (6 x 6 x 6) different messages. For this experiment, we chose five message sizes to display on the cue using five different mappings. The messages, based on one or more categories from Table 1, were mapped into the cue display using position, color, and intensity. Each cue represented information about a message that was available to the user.

Mapping 1. Here, all three lights were lit with the same high-intensity color. The color represented the source of the message; red for family, blue for friends, and green for work. This mapping used three lights to represent three messages.

Mapping 2. The three lights were the same used in Mapping 1. This time, however, the intensity of the color also varied. High intensity for a given color indicated that the message was an email. Low intensity indicated a voicemail. For example, if the lights were high intensity blue, this indicated an email message from friends. This mapping used the three lights to represent a total of six (3x2) messages to the user.

Mapping 3. The three lights were used, but each with three high-intensity colors (red, blue, green). The left light indicated source, the center type (new, reply, forwarded), and the rightmost priority (high, medium, low). Each light was lit for each notification. For example, “blue green red” indicated a forwarded message from friends with high priority. The lights represented twenty-seven (3x3x3) messages.

Mapping 4. The three lights were used as in Mapping 3. In addition, two intensity levels were used with the left light (source) to indicate medium (email, voicemail). For example, “blue (low intensity) green red” indicated a forwarded voicemail from friends with high priority. This mapping represented fifty-four (6x3x3) messages.

Mapping 5. Mappings were the same as in Mapping 4, with the addition of two intensity levels for the center light (type) to indicate length (long, short). For example, "blue (low intensity) green (high intensity) red" indicated a long forwarded voicemail from friends with high priority. This mapping represented 108 (6x6x3) messages.

These five mappings were used to create five message-sets. According to information theory, the amount of information in a message is related to the number of possible alternative messages – i.e. the more alternatives, the greater the information. Information is measured in bits – the number of binary decisions needed to identify a single message out of all the possible alternatives. This is represented by:

$$H = \log_2 N \tag{1}$$

where N is the number of alternative messages in the message-set [12]. Information loads range from 1.58 to 6.75 bits for message-sets 1 through 5 (see Table 2).

Display dimensions were chosen to produce intuitive mappings. For example, color is an especially effective way of coding nominal sets [7]. Kahneman and Henik [6] pointed out that the human visual system is effective at distinguishing a small number of distinct colors, so the three colors red, blue, and green were used. In addition, two easily discernable intensity levels were used for each color.

Table 2. Information load for each message-set

Message-Set	Alternatives	H (bits)
1	3	1.58
2	6	2.58
3	27	4.75
4	54	5.75
5	108	6.75

The message-sets are essentially notifications containing meta-data about messages for the user. These messages-sets were designed to facilitate learning and improve the performance of a three pixel display by (a) organizing messages by categories of meta-data, (b) mapping these categories directly to cues, and (c) providing for a progressive learning style. Categories provide subjects with a way to create mental "chunks" of information, thereby reducing the work of identifying messages. Without chunking, each message in the set of alternatives would have to be remembered, but with chunking, only the alternatives of a set of categories need be identified. Learning is further facilitated by providing a simple, direct, and consistent mapping between information and cues. In this case, the categories are mapped directly to visual cues. Finally, each larger message-set contains the mapping of the previous sets so that learning can be progressive, or built-up over time.

4 Methodology

Fifty-two undergraduate and graduate college students participated in this study. Eight subjects were female and forty-four were male. Ages ranged from eighteen to thirty years with an average age of twenty-three. None reported themselves as colorblind.

Subjects were required to learn each message-set to a criterion level in order from the easiest (set 1) to the most difficult (set 5). They were only allowed to advance to the next message-set after achieving 90% correct on the current set.

The experiment was then conducted on a Pentium-4 computer running Windows XP with a 1024 x 768screen resolution. A Java program presented the cue displays as Graphics Interchange Format (GIF) files on a black background with a taskbar at the bottom. The GIF files were 555x250 pixels, with 96 pixels/inch resolution and 8-bit color depth. Status indicators showed the elapsed trial time and the number of correct answers for the current session. See Figure 1 for a sample program screen.

4.1 Design

The design was a one-factor (message-set) repeated measures design with five levels (see Table 2). Dependent measures included number of trials to criterion, time to criterion, response time per trial, and first-click response time. The number of trials per block varied according to how many trials it took to reach the criterion performance level (90%). Messages were selected for presentation in random order without replacement within each block.

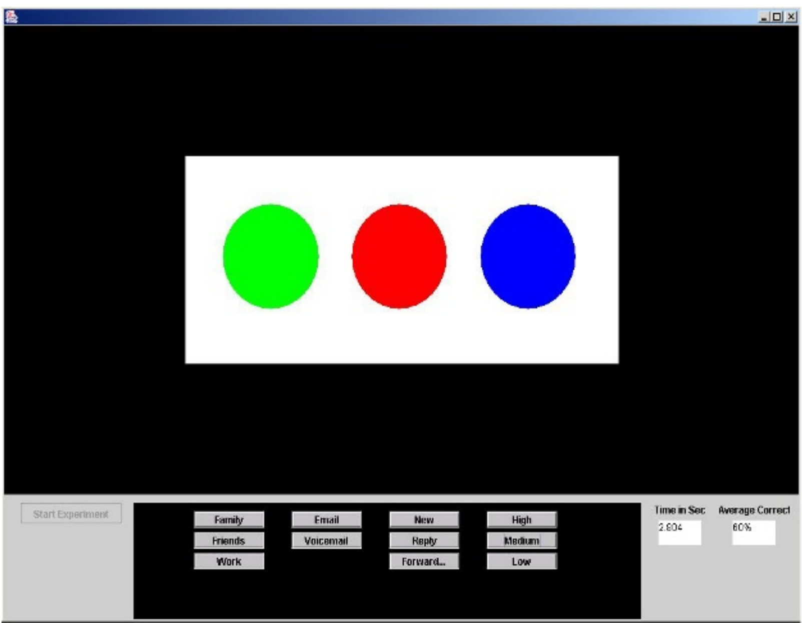


Fig. 1. Screen shot of testing environment

4.2 Procedure

Subjects first completed a questionnaire asking for background information. Each subject then completed five task sessions of increasing complexity that involved identifying notification cues. At the beginning of each session, a subject was shown a visual explanation of how information from a notification cue mapped to a specific visual display. The mappings started with Mapping 1 for the first session and finished with Mapping 5 for the last session. When a subject was ready to proceed, they were shown a notification cue. A subject responded by selecting one or more buttons on the screen corresponding to the information that was conveyed by the cue. A subject was then shown whether or not their response was correct, and they moved onto another cue in that session when they were ready. Subjects had a maximum of eight seconds to respond to each cue; otherwise, the cue timed out and was counted as incorrect. Subjects continued with a particular session until they got 90% of their responses correct, at which time they proceeded to the next session. Subjects that completed all five sessions were given US$5 (otherwise, no payment was given). Each subject proceeded with the experiment at their own pace (except for the time-out), and could stop the experiment at any point. Response accuracies and task times were recorded.

During each session, buttons with the answer choices were listed in columns at the bottom of the screen. Once a selection from each column was made, or when the question timed out, the program highlighted the correct answer on the buttons. The percentage of correct answers for each session was displayed after each response, but the first determination of whether or not to proceed to the next session was made after the first ten answers. Thereafter, the percentage was calculated after every answer on a moving basis over the ten most recent responses.

5 Results

Results were analyzed by calculating the number of trials and time to reach criterion for each set of messages. The number of trials was simply a count of trials in each condition that were performed before the running average reached 90% correct or greater. Note that accuracy and number the trials are inversely correlated; the higher

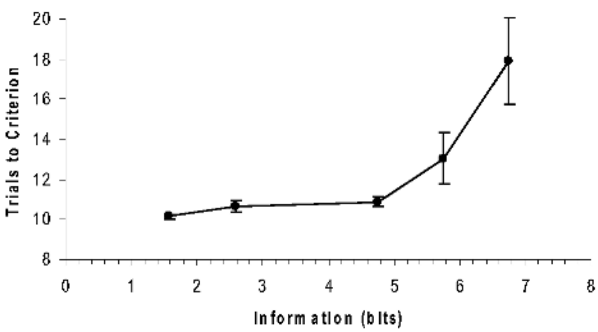


Fig. 2. Mean number of trials to reach criterion performance across information levels (±SE)

the accuracy, the fewer number of trials. Because the running average was calculated over a window of ten trials, the lowest number of trials in a condition is ten. Time to reach criterion was calculated by summing the times of all trials in a condition. As the message-set factor was within-subjects, all ANOVAs were performed using a repeated measures analysis and t-tests were performed using paired samples.

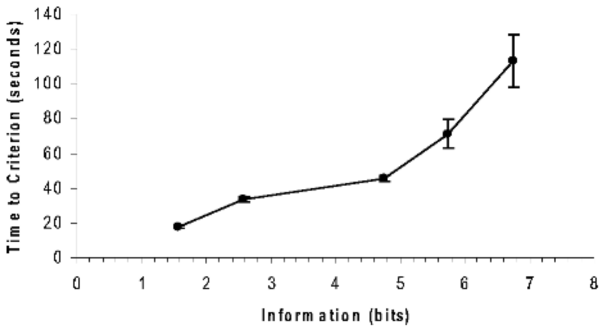


Fig. 3. Mean time to reach criterion performance across information levels (±SE)

A one-way repeated measures ANOVA showed a reliable increase in the number of trials needed to reach criterion across conditions ($F(4,180) = 8.30$, $p < 0.001$). As shown in Figure 2, the mean number of trials to criterion stays at ceiling performance (around 10 trials) for the first three message sets; but, significantly drops below the ceiling for message sets 4 (5.75 bits) and 5 (6.75 bits) ($t(45) = 3.42$, $p < 0.001$).

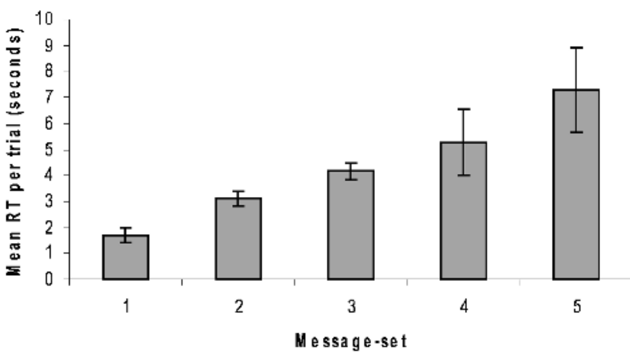


Fig. 4. Mean response time (RT) per trial across message-sets (±SE)

A one-way repeated measures ANOVA showed a reliable effect of message-set on the time to criterion ($F(4,180) = 26.04$, $p < 0.001$). As Figure 3 shows, time to criterion increases steadily across information levels from a mean of 17 seconds to almost 120 seconds. Pairwise comparisons support this by showing reliable differences between information levels at 1.58 and 2.58 bits ($t(45) = 10.7$, $p < 0.001$); 2.58 and 4.75 bits ($t(45) = 5.5$, $p < 0.001$); 4.75 and 5.75 bits ($t(45) = 3.2$, $p = 0.0015$); 5.75 and 6.75 bits ($t(45) = 2.9$, $p = 0.0032$).

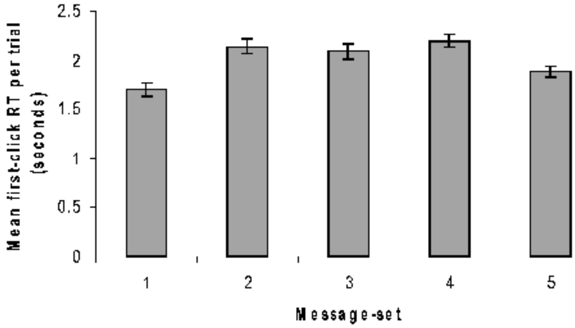


Fig. 5. Mean first-click response time (RT) per trial across message-sets (±SE)

One difficulty with calculating the total time to criterion is that the increase in time could be the result, merely, of performing more trials with bigger message-sets. To examine this possibility, the mean response time per trial was calculated and is shown in Figure 4. This shows a reliable increase in response time per trial as the message-sets increase in terms of alternatives ($F(4,180) = 502.85$, $p < 0.001$). However, the increase is not as pronounced as in Figure 3.

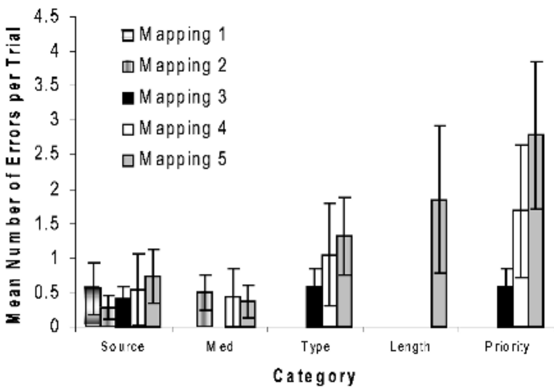


Fig. 6. Mean number of errors per trial across cue category for each mapping (±SE)

One could further argue that the increase in response time for larger message-sets is solely the result of having to select more buttons in order to respond. Therefore, the mean first-click response time was analyzed (see Figure 5). This is calculated as the time from stimulus presentation until the first button is clicked. All these times are summed until criterion performance is reached and divided by the number of trials. A one way repeated measure ANOVA indicated a reliable increase in first-click response time across message-sets ($F(4,180) = 17.16$, $p < 0.001$). Subjects typically wait 2 seconds after the notification is displayed before responding except for message-sets 1 and 5.

Analysis of errors by cue category (see Table 1) shows approximately the same number of errors being made across all relevant mappings for Source and Medium (see Figure 6). There is, however, an increase in the number of errors for Type and

Priority. This is shown as a significant increase in errors for Type and Priority over the remaining cue categories in Mappings 4 ($F(1,46) = 8.38$, $p = 0.006$) and 5 ($F(1,46) = 23.255$, $p < 0.001$).

Analysis of time-outs – the trials that timed-out due to no response before the time limit – showed that very few time-outs occurred during the experiment. Only 9 trials for all subjects combined timed out and there was no reliable difference across message-sets ($F(4,180) = 1.71$, NS).

6 Discussion

The results show very good performance by many subjects as performance was near ceiling for message-sets 1, 2, and 3. In other words, subjects had no trouble learning up to 4.75 bits of information in 10 trials or less. To learn all 6.75 bits of information or 108 alternatives required only 19 trials on average. Performance, however, does start to decline more dramatically after about 5 bits of information. Also, response times showed that this high accuracy was achieved at the expense of time. Essentially, the response times increase steadily over message-sets and significantly increase across message-sets 2 and 3 (see Figure 3). This shows that there is a cognitive cost to learning larger amounts of information from the same size display even if this cost is not reflected in the accuracy alone.

The argument could be made that these effects are due to the increased number of buttons that need to be clicked across message-sets. However, looking at first-click response times – the time from display presentation to the first click – we still see a significant increase in response time (see Figure 4). This assumes that subjects use a response strategy of first identifying all the cues and then responding. Several response strategies are possible:

1. Identify all cues and then respond.
2. Respond immediately with identified cues, identify remaining cues, then finish response.
3. Identify some cues, respond, identify more cues, respond, identify remaining cues, respond, etc.

It could be the case, for example, that when the number of response choices exceeds a certain amount, subjects switch to the second strategy listed. In other words, when faced with many buttons, subjects immediately click the buttons they know are correct, stop to consider the remaining choices, and finish responding. While this study does not definitively uncover which response strategy is used, the increasing first-click response time combined with the increasing trials to criterion suggest greater cognitive effort to identify notifications with more information.

Some support for strategies two or three comes from the analysis of errors by category (Figure 6). The number of errors should decrease for the same cue category repeated across Mappings due to learning. But subjects are probably reading and interpreting the cues from left to right, and the rightmost light receives the least attention. Even though Priority is interpreted exactly the same way in Mappings 3 and 4, Mapping 4 has an additional category to interpret on the leftmost light. This leaves less time to interpret the remaining lights because of the time limit, and results in more mistakes. This would also explain the increase in mistakes for Priority on

Mapping 5 (along with the relatively high error rate for Length), because there is now an additional category to interpret on the rightmost light.

This suggests more important information should be placed on the "leftmost" part of a pixel-based notification display when the display is composed of elements aligned horizontally. Another potentially useful result is that, at least for the leftmost light of this three light cue design, there is no significant difference in error rates of the category mapped to color (Source) or to intensity (Medium).

7 Conclusions and Future Work

The goal of this study was to investigate the amount of information that could realistically be presented on pixel-based micro-sized displays. Results indicate that people can quickly learn fairly large notifications of over six bits with only three pixels. This makes low-information-rate, micro displays practical for people not willing to endure extensive training sessions. Design possibilities are also enhanced because many message schemes can be used with over six bits of information.

Clearly, there are a number of factors that contributed to finding robust performance over increasing information rates. Among them is the fact that stimulus-response (S-R) compatibility was high. S-R compatibility is the ease of transformation between the stimulus representation and the response [11]. In our case, the mapping between message categories (i.e., source, medium, type, length, priority) and response categories (i.e., set of buttons for each category) was direct. If, on the other hand, we had presented people with an array of buttons for each alternative message, performance would have degraded more quickly.

Another important factor was the organization of the message-sets into categories or chunks. For example, the 108 messages of set 5 could be decomposed into five categories. So, instead of having to identify one out of 108 unrelated messages, the subject only needs to identify 5 categories with 2-3 alternatives per category. Chunking can also provide a variety of response strategies. As noted above, people can respond in a piecemeal fashion instead of all at once, effectively simplifying the identification task. Future work will compare message-sets that can be organized by category against those that cannot, and also against those for which the user provides a customized organization. This comparison will allow designers to determine the limitations of low-information rate displays for less structured information.

Designing hierarchical message-sets allows people to progressively learn instead of trying to memorize all messages at once. This appears to increase the size of messages that can be conveyed and reduce learning requirements. It also has the benefits of:

- Allowing for immediate use of the notification system with almost no learning
- Providing advanced functionality for advanced users while still offering simpler functionality for other users.

Future work will compare methods for designing notifications or messages-sets. It would be valuable to know, for example, if providing hierarchical information-cue mappings facilitates learning when presented in (a) progressively more difficult training blocks, (b) random blocks, or (c) all in one difficult block. Other work will look at how customized information-cue mappings affect notification ease of use. Smaller form factors (e.g., watches) and other problem domains will also be explored.

There are many practical benefits to using pixel-sized visual notification cues. Theoretically, small lights (e.g., LEDs) can be embedded in almost any device or product. The cues can be sent quietly. They can also be customized to address privacy and security concerns. For example, three blue lights on a person's ring, even when noticed by other people, could convey a message only understood by the wearer.

References

1. Hansson, R., Ljungstrand, P., and Redstöm, J. Subtle and Public Notification Cues for Mobile Devices. *Proc. UbiComp 2001*, Springer-Verlag (2001).
2. Harris, L. R. and Jenkin, M. Vision and Attention. In M. Jenkin and L. R. Harris (Eds.), *Vision and Attention*, Springer-Verlag (2001), 1-17.
3. Holland, S. and Morse, D. R. Audio GPS: Spatial Audio in a Minimal Attention Interface. *Proc. Mobile HCI 2001*.
4. Holmquist, L. E., Falk, J. and Wigström, J. Supporting Group Collaboration with Inter-Personal Awareness Devices. *Personal Technologies*, 3(1-2), 13-21, 1999.
5. Johnson, P. Usability and Mobility; Interactions on the Move. *Proc. 1st Workshop on HCI for Mobile Devices*, Univ. of Glasgow (1998).
6. Kahneman, D., and Henik, A. Perceptual Organization and Attention. In M. Kubovy and J. R. Pomerantz (eds.) *Perceptual Organization,* Lawrence Erlbaum (1981), 181-211.
7. Mackinlay, J. Automating the Design of Graphical Presentations of Relational Information. *ACM Trans. on Graphics*, 5(2), 110-140, 1986.
8. Maglio, P. P. and Campbell, C. S. Tradeoffs in Displaying Peripheral Information. *Proc. CHI 2000*, ACM Press (2000), 241-248.
9. McCrickard, D. S., and Chewar, C. M. User Goals and Attention Costs. *Comm. of the ACM*, 46(3), 67-72, 2003.
10. McCrickard, D. S., Catrambone, R., Chewar, C. M., and Stasko, J. T. Establishing Tradeoffs that Leverage Attention for Utility: Empirically Evaluating Information Display in Notification Systems. *Intl. Journal of Human-Computer Studies*, 8, 547-582, 2003.
11. Morin, R. E. and Grant, D. A. Learning and Performance of a Key-Pressing Task as a Function of the Degree of Spatial Stimulus-Response Correspondence, *Journal of Experimental Psychology*, 49, 39-47, 1955.
12. Shannon, C. E. & Weaver, W. *The Mathematical Theory of Communication*. University of Illinois Press (1949).
13. Tarasewich, P. Designing Mobile Commerce Applications. *Comm. of the ACM*, 46(12), 57-50, 2003.
14. Tarasewich, P., Campbell, C., Xia, T., and Dideles, M. Evaluation of Visual Notification Cues for Ubiquitous Computing. *Proc. UbiComp 2003*, Springer-Verlag (2003), 349-366.
15. Wisneski, C. A. *The Design of Personal Ambient Displays*. M.Sc. Thesis, May 1999, MIT Media Laboratory, Cambridge, MA.

Investigating the Relationship Between Battery Life and User Acceptance of Dynamic, Energy-Aware Interfaces on Handhelds

Lance Bloom[1], Rachel Eardley[2], Erik Geelhoed[2], Meera Manahan[1], and Parthasarathy Ranganathan[3]

[1] Hewlett Packard, USA
{lance.bloom, meera.manahan}@hp.com
[2] Hewlett Packard/HP Labs, UK
{rachel.eardley, erik.geelhoed}@hp.com
[3] Hewlett Packard/HP Labs, USA
partha.ranganathan@hp.com

Abstract. In a 24 x 7 mobile world experiencing a proliferation of handheld devices, battery life can be a limiting factor. In particular, handheld displays consume substantial battery power. One strategy to potentially reduce display battery consumption and support a positive user experience is to adopt emerging display technologies (e.g., OLEDs) that support energy-aware interfaces. The research reported here, the second investigation in a series, assessed user expectations regarding handheld battery life and explored the relationship between battery life and user acceptance of energy-aware, handheld interfaces. Twelve experienced handheld users engaged dynamic, prototype energy-aware interfaces to complete a scenario comprised of 5 representative tasks. Users identified battery life as an important handheld issue, were positive regarding a display-based approach to reducing battery consumption and varied consistently in their enthusiasm for specific interfaces. The findings highlight themes for the research and design of future energy-aware interfaces.

1 Introduction

With their ongoing proliferation and evolution, mobile handheld devices (e.g., MP3 players, PDA's, Cell Phones and Smart Phones) continue to represent an important hardware, software and services market [2]. However, battery life can fundamentally limit the functional utility of these devices [6]. In particular, handheld displays consume substantial energy [6, 7] that can account for nearly 60% of the total system power consumption [1]. Moreover, unlike other system components, display power consumption traditionally has remained relatively constant as devices become smaller. Thus, display power consumption may represent an increasingly large proportion of the total system power consumption of future smaller devices. Contemporary strategies for reducing handheld display power consumption include powering off the device following a pre-defined interval of nonuse, designing devices with small displays and designing devices with reduced-quality displays. An alternative strategy for reducing display power consumption is to adopt emerging display technologies. One such technology, Organic Light Emitting Diodes (OLEDs), can reduce display battery consumption [5] by enabling *energy-adaptive interfaces*

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 13–24, 2004.

that consume energy only from specific regions of a display, such as those relevant to the user task. Therefore, these energy-adaptive interfaces have the potential to simultaneously reduce display power consumption and provide a positive user experience. Indeed, energy-adaptive interfaces have been found to reduce battery consumption up to a factor of 10 in laptop computers [4]. However, it is not clear how the adaptive nature of these interfaces – which can dynamically modify the brightness, color and power status of display regions – impact the user experience. We appear to be the first to investigate the relationship between battery life and user acceptance of dynamic, energy-aware interfaces on handhelds.

The goal of our research was to assess user expectations regarding mobile handheld battery life, and explore the relationship between battery life and user acceptance of dynamic, energy-adaptive handheld interfaces. From our research, we endeavor to: identify battery life parameters that are acceptable to users; understand the relationships (e.g., tradeoffs, enhancements, etc.) between the energy-saving and user-acceptance aspects of energy-aware interface designs; distill specific designs and design principles that maximize battery life and user acceptance of future energy-aware interfaces; and, identify user tasks and applications that can potentially benefit from these designs and design principles. In our first investigation [3], we found that participants were generally accepting of energy-aware interfaces, particularly notification and menu interfaces with high-contrast areas that promoted the interface region salient to the user task. The investigation reported here, the second in the series, went beyond the first and made several unique contributions that included: sharpening the focus of investigation regarding user expectations of handheld battery life; broadening the scope of evaluation to include a new and wider variety of interfaces and software applications; displaying dynamic interfaces on a PDA; and, recruiting participants not employed by our company and who all did not own the handheld brand sold by our company. These unique contributions served to enhance the scope and validity of our findings.

2 Methods

In this section we describe our user-evaluation methodology. To summarize, 12 experienced handheld users engaged dynamic, prototype energy-aware interfaces on a handheld device and completed a scenario comprised of 5 representative tasks. For each of 5 interface types, users evaluated multiple interfaces including a 'control' interface in contemporary use. Each interface displayed a unique combination of visual appearance and battery life. Based upon the battery life, visual appearance and perceived usability of each interface, users provided ratings, verbal comments and direct-comparison data.

2.1 Participants

Participants were 12 experienced PDA users from the Boston area that were contacted through a market research firm. The representative sample of users included men and women as well as a range of occupations (e.g., VP, Sales Rep., Teacher, Engineer,

Owner, Manager) and industries (e.g., Financial, Healthcare, Retail, Consulting, Gov't., Hi Tech) from small-, medium- and enterprise-scale employers. All participants regularly engaged their PDA for a combination of work and personal activities. All participants owned a PDA equipped with the MS Pocket PC OS, and most owned an iPAQ.

2.2 Materials

The dynamic, prototype energy-aware interfaces were implemented in Flash and displayed on an iPAQ h5550. Battery life was specified in an icon at the top, right of the interface (e.g., '4h 39m' indicated 4 hours and 39 minutes). The battery-life estimates were derived from an engineering power analysis of the prototype interfaces. Prototypes supported user scrolling and the MP3 interfaces were fully functional. Five types of dynamic interfaces were displayed: E-mail inbox gradient interfaces; Acrobat Reader gradient interfaces; MP3 player interfaces; inversion interfaces; and, flashlight interfaces. Examples of the dynamic interfaces used in this investigation, and their respective battery lives, are displayed in Appendix A of this document.

Gradient Interfaces. Participants viewed 6 gradient interfaces (0, 20, 40 60, 80 & 100%) for both the e-mail inbox and reader applications. The 0% interfaces were not gradients, but rather the conventional MS Outlook Inbox and Acrobat Reader interfaces, and they served as comparison (control) interfaces in the present investigation. Compared to the contemporary (control) interfaces, the energy-aware interfaces achieved energy reductions of up to a factor of 2.5 for the inbox and up to a factor of 6 for the reader.

MP3 Interfaces. Participants viewed 3 MP3 interfaces: the contemporary (control), blue windows media player; a gray and black interface; and, a green and black interface. Compared to the contemporary (control) interface, the energy-aware interfaces achieved energy reductions of up to a factor of 21.

Inversion and Flashlight Interfaces. Participants viewed 4 inversion interfaces (start, calendar, e-mail inbox and Acrobat Reader) that displayed black backgrounds with white text -- except for the start interface, which displayed a tan background with black text. Participants also viewed several flashlight interfaces that displayed a dimmed interface with a user-movable region that was illuminated at standard levels. These interfaces enabled participants to move the 'flashlight' by depressing the stylus on the illuminated area (e.g., the edge) and dragging it. Several versions of the flashlight interfaces were created by varying 2 dimensions: the color of the dimmed area (gray, black); and, the shape of the illuminated area (square, horizontal rectangle). Compared to the contemporary (control) interfaces, the energy-aware interfaces achieved energy reductions of up to a factor of 5 for the inversion interfaces and up to a factor of 9 for the flashlight interfaces.

2.3 Design

Each session was comprised of 4 data-collection components completed in the following sequence: The participant background and handheld usage data-collection component; the gradient interface data-collection component (e-mail and reader); the MP3 interface data-collection component; and, the inversion and flashlight interface data-collection component.

Within each data-collection component, the specific interface presentation sequence was counterbalanced across participants to eliminate uninteresting interpretations of the data. Thus:

The gradient and interface-type presentation sequences were counterbalanced orthogonally. Specifically, half of the participants viewed the gradients in sequence from 0 to 100% and half viewed the gradients in sequence from 100 to 0%. Half of the participants viewed the e-mail interfaces prior to the reader interfaces and half viewed the reader interfaces prior to the e-mail interfaces.

A Latin square was used to counterbalance the presentation sequence of MP3 interfaces.

The inversion and flashlight interface presentation sequence was counterbalanced such that, half of the participants viewed the inversion interfaces prior to the flashlight interfaces and half viewed the flashlight interfaces prior to the inversion interfaces. The presentation sequences of the individual inversion and flashlight interfaces were also counterbalanced.

2.4 Procedure

The evaluation was conducted in a typical, well-illuminated office environment. One individual participated in each 90-minute session. Each session began with the participant providing background information regarding their PDA usage, observed PDA battery life and desired PDA battery life. The participant then performed 5 tasks as part of a scenario in which s/he traveled by train to meet with a business customer. Specifically, the participant reviewed an e-mail inbox and read a page from a book during the train ride to the customer meeting, viewed and used the MP3 player on the return train journey and, also on the return journey, viewed each of 4 inverted interfaces and multiple flashlight interfaces to reduce display consumption of dwindling battery power.

For each individual interface, participants engaged the prototype to complete the task (e.g., scroll and read), offered verbal remarks and then provided ratings based upon battery life, interface appearance and usability. After participants viewed all interfaces of one type (e.g., all 6 e-mail gradient interfaces), they completed a direct-comparison task based upon battery-life, visual-appearance and usability criteria. Specifically, for the gradient interfaces, participants specified the 1 interface of that type that they were most likely to use. For the MP3 interfaces, participants rank ordered the 3 interfaces to indicate their 1st through 3rd choices. And for the inversion and flashlight

interfaces, participants were given the option to choose the inversion interfaces, the flashlight interfaces or indicate no preference.

3 Results

In this section we present the data regarding participant expectations of handheld battery life, and participant acceptance of handheld, energy-aware interfaces. To summarize, participants indeed indicated that handheld battery life is an important issue, and that they expected a longer battery life than currently supported by their device. In general, they were favorable regarding a display-based approach to reducing battery consumption and indicated that they wanted a choice of display settings, such that each setting provided a unique combination of battery life and interface visual appearance. Regarding specific interface types, participants were quite positive towards the inversion and MP3 interfaces and less positive towards the gradient and flashlight interfaces.

3.1 Battery-Life Expectations

PDA battery life was an important issue for participants and they were receptive to a display-based approach to extending battery life.

Participants rated the importance of several PDA attributes on a scale ranging from +2 (important) to 0 (neutral) to -2 (not important). Battery life received the third highest (i.e., most important) mean rating (+1.67). The ratings were obtained during participant recruiting, prior to their knowledge of the content or purpose of the study. The means, displayed in Table 1, are sequenced from most to least important.

Table 1. PDA attribute rating data

Memory	Processor Speed	Battery Charge Life	Productivity Applications	Screen Size	Graphics	Size	Communication Applications
1.75	1.75	1.67	1.33	1.17	1.08	1.08	1.08

During the testing sessions, several participants stated that battery life was an important issue. Most participants indicated that their PDA provided a battery life of 2 to 4 hours with continuous usage, and they all desired a longer battery life, with the most common request being a two- to three-fold proportional increase or an 8-hour absolute battery life with continuous usage. These requests were made to support a usage model characterized by a full day of work followed by recharging at night; most participants indicated that they did not want to carry additional equipment (e.g.,

charger, extra battery) on a daily basis. Several participants requested the option of choosing from display settings, such that each setting provided a unique combination of battery life (e.g., 2, 4, 6, 8 hours) and interface visual appearance.

Participants liked the conspicuous presentation of the remaining battery power at the top of the interfaces. Users were aware that their PDA contains a power-settings interface, but they would prefer to have the information conspicuous at all times, because they tend not to navigate to view the information and therefore typically do not know how much battery life remains. As one participant stated, "so if I was working for 15 minutes I could see how much battery power I'm using."

3.2 Gradient Interfaces

Ratings, direct-comparison data and verbal comments consistently indicated that participants were less than enthusiastic about the gradient interfaces because they were confusing, particularly if the participants scrolled, and because the interfaces did not facilitate the tasks of scanning and reading. Some participants stated that they would prefer to extend battery life by selecting from preexisting display settings, such that each setting provided a unique combination of battery life and a single-color (gray) interface background.

Ratings Data. After viewing each interface, participants used a scale ranging from +4 (definitely yes) to 0 (neutral) to -4 (definitely no) to rate the following statement: 'Overall, I would use this interface design on a regular basis.' Participant rating data for the gradient interfaces is displayed in Table 2. Overall mean rating decreased as the gradient % increased (i.e., became darker), $F(5, 55) = 5.91$, $MSE = 34.73$, $p = .004$. This finding also was observed for the respective e-mail and reader ratings, both p's < .05. Pairwise comparisons indicated that, for both the e-mail and reader interfaces, only the darkest gradient (100%) was rated lower than the control interface (0%), both p's < .05. Finally, the e-mail interface mean rating (1.42) was not reliably higher than the reader interface mean rating (0.97), $F(1, 11) = 2.04$, $MSE = 7.11$, $p = .18$.

Table 2. Gradient rating data

	Gradient						
Interface Type	**0%**	**20%**	**40%**	**60%**	**80%**	**100%**	**Mean**
E-mail	2.17	1.92	1.92	1.58	1.08	-0.17	**1.42**
Reader	1.75	1.75	1.33	1.17	0.42	-0.58	**0.97**
Mean	**1.96**	**1.84**	**1.63**	**1.38**	**0.75**	**-0.38**	**1.20**

Direct-Comparison Data. After rating all 6 gradients of each interface type, participants indicated which of the 6 gradients they would most likely use, based upon battery life, interface visual appearance and usability. The results are displayed in Table 3. Consistent with the ratings data, there was some participant interest in the 40% and 20% gradient screens. However, for both the e-mail and reader interfaces, the contemporary or control (0% gradient) interface was chosen most frequently. Two CHI Square tests, each with 6 categories, indicated that these findings differed from

chance. For the e-mail gradients, $X^2(5) = 15$, $p = .01$, and for the reader gradients, $X^2(5) = 18$, $p = .003$.

Table 3. Number of participants who chose each gradient

	Gradient					
Interface Type	**0%**	**20%**	**40%**	**60%**	**80%**	**100%**
E-mail	6	1	4	1	0	0
Reader	7	1	3	0	0	1

Verbal-Comment Data. Participant ratings were, on average, less favorable for the 100% gradient than the 0% gradient because participants 'could not see the entire screen.' Additionally, participants stated that all of the gradients were somewhat distracting or confusing, particularly if they scrolled. Regarding the e-mail inbox, several participants commented that it was not clear which e-mail was highlighted. For the Acrobat Reader, several participants stated that the gradients imposed a discrete, artificial window on a continuous process that required them to see the entire page. Several participants noted that they would prefer inbox and reader interfaces with a single-color, light-gray background that would reduce battery consumption and enable sufficient contrast with superimposed text so that all of the text on the interface was easy to scan or read, depending upon the task. For example, the participants often commented that they liked the darkest shade of gray on the 40% gradients as a candidate for a single-color, gray background.

3.3 Inversion and Flashlight Interfaces

Ratings, direct-comparison data and verbal comments consistently indicated that participants liked the energy-aware inversion interfaces, tended to prefer them to the contemporary (control) interfaces and clearly preferred them to the flashlight interfaces. The inversion interfaces were received favorably because they reduced battery consumption relative to the contemporary interfaces and were readable. Participants generally perceived the flashlight interfaces as novel but lacking a task application.

Ratings and Direct-Comparison Data. After viewing each interface, participants used a scale ranging from +4 (definitely yes) to 0 (neutral) to -4 (definitely no) to rate the following statement: 'Overall, I would use this interface design on a regular basis.' The mean rating for the inversion interfaces (+3.08) was more favorable than the mean rating for the flashlight interfaces (0.00), $t(1, 11) = 3.56$, $p = .004$. Moreover, the mean rating for the inversion interfaces (+3.08) tended to be more favorable than the mean rating for the contemporary (control) interfaces (+1.96), $t(1, 11) = 2.08$, $p = .06$. After rating the inversion and flashlight interfaces, participants performed a direct-comparison task. Based upon battery life, interface visual appearance and usability, 9 of the 12 participants preferred the inversion interfaces, 2 participants expressed no preference and 1 preferred the flashlight interfaces. A CHI Square test with 3 categories indicated that these findings differed from chance, $X^2(2) = 9.5$, $p = .009$.

Verbal-Comment Data. Participants generally perceived the flashlight as a novelty without any practical application, despite each participant and the facilitator identifying several potentially-relevant tasks and scenarios. All 12 participants stated that they liked the inversion interfaces and would use them in scenarios in which reducing battery consumption was at issue. Several participants further indicated that they would consider using the inversion function as their default setting on a trial basis. Participants liked these interfaces because they provided substantial power savings compared to the contemporary (control) interfaces and they were easy to read.

Participants also noted that the implementation of the inversion interfaces was important. Thus, although several participants commented positively on the strong contrast afforded by white text on a black background (e.g., Acrobat Reader Interface), a few participants indicated that the text was a bit small or spindly, and therefore somewhat difficult to read. These participants wanted to select font size and type. Some participants also noted that they did not like the specific colors implemented in the inverted start interface.

3.4 MP3 Interfaces

Ratings, direct-comparison data and verbal comments indicated that, overall, the gray interface was the most popular, the green interface was the second-most popular, and the blue interface was the least popular. The gray and green interfaces were received favorably because they reduced battery consumption relative to the blue (control) interface, and they also provided a good visual design.

Ratings and Direct-Comparison Data. After viewing each interface, participants used a scale ranging from +4 (definitely yes) to 0 (neutral) to -4 (definitely no) to rate the following statement: 'Overall, I would use this interface design on a regular basis.' The gray MP3 interface received the most favorable mean rating (+2.08), followed by the green interface (2.00) and the contemporary (control) blue interface received the lowest mean rating (1.50). However, these differences were attributable to chance, $F < 1$.

After rating the interfaces, participants rank ordered the 3 MP3 interfaces to indicate their first (1) through third (3) choices. For their first choice, 6 participants chose the gray interface and 6 chose the green interface. A CHI Square test with 3 categories showed that these findings differed from chance, $X^2(2) = 6$, $p = .05$. The overall mean ranking for the three interfaces indicated that the gray interface was ranked most favorably (1.58), followed by the green interface (1.83) and finally the blue interface (2.58), $F(2, 11) = 4.09$, $MSE = 4.78$, $p = .05$. Comparison of means revealed that the mean ranking for the gray interface (1.58) was superior to the mean ranking for the contemporary (control) blue interface (2.58), $p < .05$.

Verbal-Comment Data. Participants preferred the gray and green interfaces to the blue interface largely because of their power-saving ability. For example, one participant who chose the green interface stated, "It's an mp3 player. I program it and stick it in my pocket. I don't look at it much." However, some participants also preferred the visual design of the gray and the green MP3 players relative to the blue

player. Finally, some participants wanted the option of selecting from preexisting MP3 display settings (skins), such that each setting provided a unique combination of battery life and interface color/illumination.

4 Conclusions

As handhelds proliferate and evolve it becomes increasingly important to find new strategies to address --what one author recently called the handheld "Achilles Heel" [6] -- their battery life. Handhelds that reduce display battery consumption have been developed, but they often invoke a sleep mode, reduce the size of the display or reduce the quality of the display and thereby risk degrading the user experience. Alternatively, emerging display technologies (e.g., OLEDs) that enable energy-adaptive interfaces can potentially reduce display battery consumption and promote a positive user experience. Recent findings do indicate that energy-aware interfaces can greatly reduce battery consumption. However, it is not clear how these interfaces impact the user experience. We appear to be the first to investigate the relationship between battery life and user acceptance of dynamic, energy-aware interfaces on handhelds.

The goal of our research was to assess user expectations regarding mobile handheld battery life, and explore the relationship between battery life and user acceptance of dynamic, energy-adaptive handheld interfaces. Twelve experienced handheld users engaged functioning, prototype energy-aware interfaces on a handheld device in the service of completing a scenario comprised of 5 representative tasks. Based upon the battery life, visual appearance and perceived usability of each interface, participants provided ratings, verbal comments and direct-comparison data. Compared to contemporary (control) interfaces, the energy-aware interfaces generally achieved energy reductions of up to a factor of 4, and as much as a factor of 21.

The high-level findings that we presently observed were generally consistent with those of our previous investigation. For example, participants presently identified limited battery life as an important issue, were supportive of a display-based approach to reducing battery consumption and varied consistently in their enthusiasm for specific interfaces. That these findings presently were obtained with a new and wider variety of dynamic interfaces that were displayed on a PDA to participants not employed by our company and who all did not own the handheld brand sold by our company, all serve to increase the scope and validity of our findings.
Moreover, we presently observed some novel and, in one instance, unexpected findings. First, participants expected a longer battery life than currently supported by their handheld device, and they typically requested a two- to three-fold proportional increase relative to their current device battery life so that they could confidently complete a full day of work with their device and recharge it at night. Second, participants requested a choice of display settings, such that each setting provided a unique combination of battery life and interface appearance. Third, participants were quite favorable towards the energy-aware inversion and MP3 interfaces, and tended to prefer them to the respective contemporary (control) interfaces. Participants preferred these energy-aware interfaces because battery consumption was greatly reduced

relative to the contemporary interfaces and participants could easily view all of the text to complete their task. Fourth, participants were less favorable regarding the energy-aware flashlight and gradient interfaces. This latter finding was somewhat surprising based upon the results of our first investigation in which participants stated that they would be interested in using gradient-type interfaces. However, in that study, participant comments were based upon an informal viewing of a static, gray gradient interface at the end of the session. Thus, the different reactions expressed by participants in the two investigations likely underscores the importance of conducting formal evaluations with functioning prototypes that are displayed on handhelds to a representative sample of external participants. Fifth and finally, participants in the present investigation stated that they would be relatively more interested in using e-mail and reader interfaces with a single-color background (e.g., gray) if the background reduced battery consumption and provided sufficient contrast to render the text easily readable.

From these findings, three themes emerge that are particularly worthy of further investigation: the identification and refinement of interface design principles that support reduced display battery consumption and a positive or enhanced user experience; the assessment of interfaces with single-color backgrounds that reduce battery consumption and provide a positive user experience, particularly in the context of text-intensive interfaces; and, the evaluation of handheld personalization, including the assessment of preexisting display settings, such that each setting provides a unique combination of battery consumption and interface color/illumination. Investigating these themes using formative, prototype design and testing will likely facilitate the identification of a sufficient number, variety and quality of interface designs so that it will become meaningful to perform a summative usability evaluation measuring behavioral performance as users engage fully functional interfaces.

In summary, given the ubiquity of handheld devices, user desire for longer battery life and user desire for robust displays, we believe that new strategies are needed to address the fundamental design tension currently existing between handheld display battery consumption and the user experience. One such strategy is to utilize emerging display technologies (e.g., OLEDs) that support energy-adaptive interfaces capable of reducing display battery consumption and providing a positive or even enhanced user experience. The present findings, together with other recent data, suggest that energy-adaptive interfaces have the potential to meet these criteria. We view this as an auspicious beginning for interfaces that promise to become an important component in mobile system design.

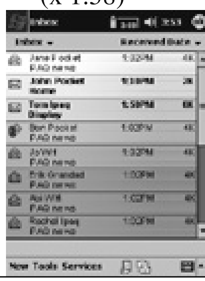
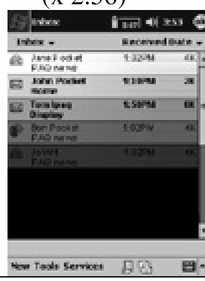
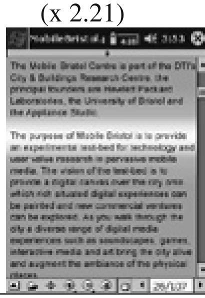
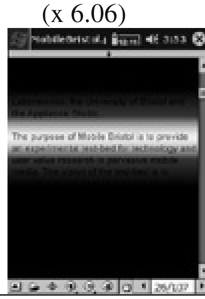
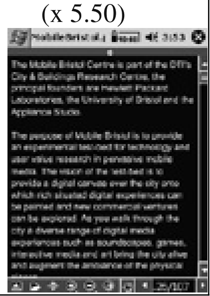
References

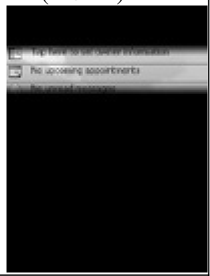
1. Choi, I., Shim, H. Chang, N.: Low-power color TFT LCD Display for Handheld Embedded Systems. Proceedings of the International Symposium on Low Power Electronics and Devices. ACM Press, Monterey (2002) 112-117
2. eTForecasts: Worldwide PDA & Smartphone Forecast 1998 – 2008 (June, 2003 Release) http://www.etforecasts.com/products/ES_pdas2003.htm

3. Harter, T., Vroegindeweij, S., Geelhoed, E., Manahan, M., Ranganathan, P.: Energy-aware User Interfaces: An Evaluation of User Acceptance. Proceedings of the Annual Conference on Human Factors in Computing Systems (2004)
4. Iyer, S., Luo, L., Mayo, R., Ranganathan, P.: Energy-Adaptive Display System Designs for Future Mobile Environments. Proceedings of the First International Conference on Mobile Systems, Applications and Systems. ACM Press, San Francisco (2003) 245-258
5. Stanford Resources Inc. (ed.): Organic Light-Emitting Diode Displays: Annual Display Industry Report, Second Edition (2001)
6. Viredaz, M., Brakmo, L., Hamburgen, W.: Energy Management of Handheld Devices. Queue. ACM Press, New York (2003) 44-52
7. Viredaz, M., Wallach, D.: Power evaluation of a handheld computer. IEEE Micro. (2003) 66-74

Appendix

Battery life is displayed relative to the baseline (control) battery life. For example, 'x 2.86' indicates a battery life that is longer than the baseline by a factor of 2.86.

Example Inbox & Reader Interfaces with Relative Battery Life			
0% Gradient	40% Gradient	100% gradient	Inversion
(baseline) 	(x 1.58) 	(x 2.56) 	(x 3.21) 
(baseline) 	(x 2.21) 	(x 6.06) 	(x 5.50) 

MP3 Interfaces with Relative Battery Life			**Flashlight**
Blue	Gray	Green	Black, Rectangle
(baseline) 	(x 2.86) 	(x 20.96) 	(x 8.77) 

Mental Models of a Cellular Phone Menu. Comparing Older and Younger Novice Users

Martina Ziefle and Susanne Bay

Department of Psychology, RWTH Aachen University,
Jaegerstrasse 17-19, 52056 Aachen, Germany
{martina.ziefle, susanne.bay}@psych.rwth-aachen.de

Abstract. The interrelationship between mental models of a cellular phone menu and performance depending on users' age was under study. The mental representation was assessed through card-sorting technique in 32 novice users (16 aged 20-32, 16 50-64 years). First, they had to process four common tasks on two simulated mobiles enabling online logging of users' actions. None of the older participants had a correct mental representation of the route to be taken to solve a task, and some were not even aware of the hierarchical nature of the phone menu. Younger participants, in contrast, had a fairly correct mental model. Furthermore, it was shown that the better the mental map of the menu, the better the performance using the device. In conclusion, the awareness of the hierarchical structure of the menu is of central importance to use a cellular phone properly. Therefore, it should be made more transparent to the user.

1 Introduction

Why do older adults in particular face extreme difficulties when starting to use a new electronic device, for example a mobile phone? As reported by Maguire and Osman [7], the development of mobile phone technology seems to concentrate on what young and experienced users want, as thrilling gadgetry, possibly because they are the most heeded user group. In contrast, for the older users, an easy to use menu is the most important issue [7]. When older people purchase their first cellular phone they are offered a number of attractive services and features which they are indeed interested in. For example they see the practicality of having a calendar, an alarm clock and a phone all in one, and having the opportunity to check train departure times while being on the move seems attractive to them. However, after getting the phone and trying to use it for a short while, these ambitious plans often end in frustration.

Disorientation is not restricted to older users. Younger people also experience difficulties with new mobile devices, as a young computer scientist states in a letter to the Economist (December 12th 2002):

"I recently bought a sophisticated mobile phone. I spent several hours trying to navigate its features and configure it to read my e-mail. I have a PhD in computer science but I still had a sense of helplessness. The battle for domination will be won

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 25–37, 2004.

not just through sleek technical innovation but by companies who consider seriously the human perspective in their designs." Matt Jones, New Zealand.

Older adults' difficulties, however, seem to be located at a more fundamental level. For an insight into their specific problems, differences between younger and older people have to be considered regarding the basal cognitive requirements for the handling of a hierarchical information structure such as the cellular phone menu.

1.1 Differences in Information Processing Between Younger and Older Adults

What are the differences in younger and older people's information processing? Several fundamental abilities necessary for information processing come into play when using a technical device with a complex hierarchical menu structure.

The decline in memory capacity in older adults is a well-known issue, as well as the decrease in the speed of processing [e.g. 4] or a reduction in resources for information processing [9]. Good memory abilities should be of central importance for the use of devices with only a small display since the user has to memorize the functions and their location within the menu. Indeed, in a recent study it was shown that the higher users' memory capacity, the better their performance using a cellular phone [3].

Further, spatial abilities have proven to decrease over the lifespan. For example a decline in mental rotation ability of 96% was shown in a study comparing 19 to 27 year olds with 66 to 77 year old participants [5]. Spatial abilities may play a substantial role for the use of cellular phones, since the menu of the phone is organized in a tree structure and spatial visualization abilities could be necessary for proper use of the menu because its functions are organized in various levels. Vicente, Hayes and Williges [12] showed in their experiment that spatial ability is of great importance for efficiency in finding information in a hierarchical arrangement of files. The positive impact of high spatial abilities on users performance navigating the functions of a cellular phones was also revealed in a recent study [3]. Zaphiris [13] has demonstrated that older adults experience particular difficulties with deep menu structures and tend to get more easily lost in broad or deep menus than young people. In the mobile phone, where the overall structure of the menu is not transparent, spatial abilities may be even more crucial, because the user has to build a mental representation of the structure when navigating through the functions. Here, older users should be even more disadvantaged than on a microcomputer task. That older people experience difficulties navigating in the user interface of a mobile was already reported in a survey [11] and demonstrated in experimental studies [3].

If navigation in a cellular phone menu can be compared to navigation in the natural environment, according to theory [10] three types of knowledge should be of importance: Landmark knowledge representing salient features on the route, procedural knowledge (or route knowledge) of the sequence of actions required to get from one point to another, and survey knowledge which represents the overall structure of the information and an overview of locations and routes in the environment. It is to be investigated whether older adults show more difficulty acquiring all three types of knowledge important for spatial orientation compared to younger users.

A third aspect which probably plays an important role is the fact that young adults today have had contact with menu-driven technology from much earlier on (e.g. video games), and this should have influenced their mental model of the functioning of menus in general – namely the tree-like structure. This knowledge can be transferred to new devices, for example cellular phones. Older adults, on the other hand, may not have a proper mental representation of a cellular phone's organization of functions in categories within a "menu" (already the concept of "menu" may be unfamiliar to them).

The present study aims at exploring the mental model of a cellular phone menu in younger and older novice users after having processed four common phone tasks on the device.

2 Method

Sixteen university students (aged 20 to 32) were recruited for the experiment. For a comparable sample of older adults with regard to educational background, 16 persons with a university degree aged 50 to 64 were selected.

The subjects processed four tasks which correspond to frequently used functions of a mobile phone: calling someone using the internal phone directory, sending a text message to a person whose number is saved in the phone directory, setting the phone to the status where the user's own number is not transmitted when calling someone and editing an entry in the phone directory. The tasks were not processed on real cellular phones but rather two models, the Nokia 3210 and the Siemens C35i, were simulated on a personal computer with a touchscreen in order to log user actions online (see Figure 1, left). Furthermore, the simulation enabled us to increase the size of the display and the keys to make sure that older participants are not disadvantaged due to poor readability of the menu or their inferior fine-motor abilities. Both simulated phones had comparable sizes (display, keys and fonts) and three menu items could be seen at a time on the display, as it is often the case in real cellular phones. A time limit of 10 minutes per task was set.

Half of the participants (8 of the younger and 8 of the older group) solved the tasks using the Nokia 3210 simulation, half using the Siemens C35i. We have chosen to use two different widely-used cellular phone models, which dispose of a comparable functionality, for reasons of ecological validity. In the following, they are not further differentiated but results are reported for both phones taken together.

Before processing the tasks on the simulated cellular phones, participants completed a questionnaire assessing age, profession and their experience using a number of technical devices (frequency of use and experienced ease using it). 21 of the 32 participants did not possess a cellular phone of their own. Of the 11 people who were owners of such a technical device, only one reported using the internal phone directory and sending text messages, while the others used it only to make and answer calls. The questionnaire was presented to the participants on a touchscreen, which enabled the users to get familiar with the experimental apparatus.

After working on the solution of the fourth task, participants were asked for their experienced ease using the mobile phone and their difficulty understanding the menu functions as well as the keys of the phone.

Then, the users' mental representation of the cellular phone's menu was assessed through a card sorting technique (see Figure 1, right).

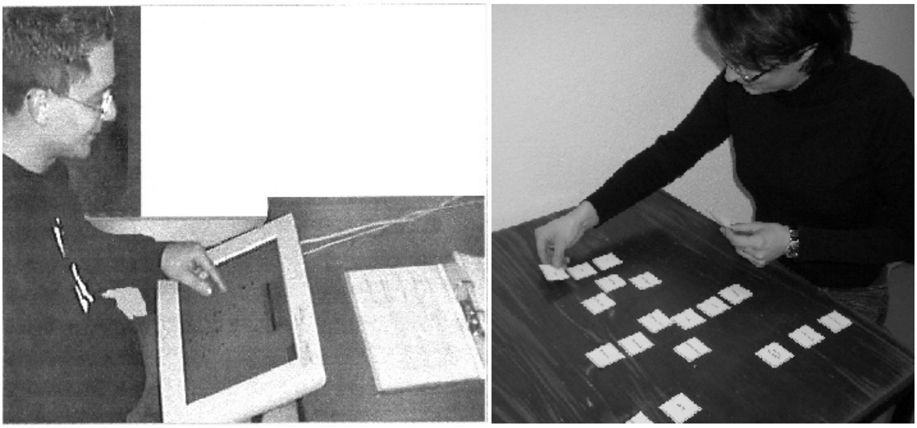


Fig. 1. Left: Participant solving the phone task on a computer simulated cellular phone; right: participant arranging the menu functions in the card sorting task

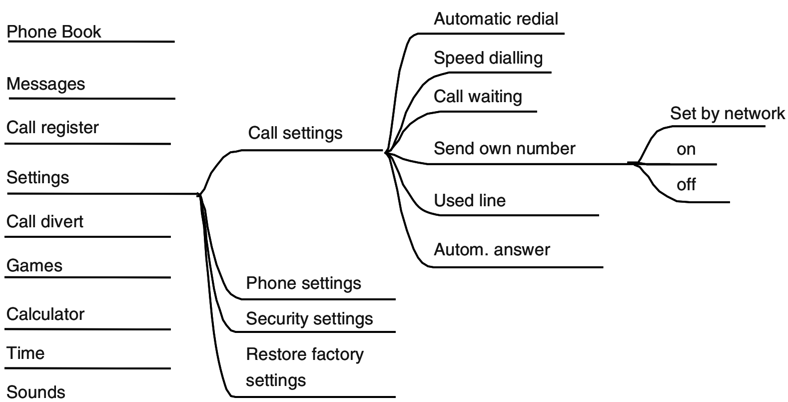


Fig. 2. Menu branch of the Nokia 3210 as example of the structure to be laid in the card sorting task

As the whole menu of the phone contains too many functions, only the menu branch where the setting of hiding the own phone number when calling is located was selected for the card sorting task. Twenty-two cards with a menu function written on each were randomly spread on the table. The 22 functions corresponded in both phones to the original items on the first level of the menu and the branch used when

setting the phone to the status where the phone number is not transmitted. The participants were asked to arrange the cards on the table according to how they remember having seen them in the menu or, if they did not remember, how it makes most sense to them. When they had finished arranging the cards, the experimenter asked the participant to explain the laid structure in a few words. Figure 2 visualizes the menu branch to be reconstructed (exemplary for the Nokia users).

3 Results

The results section will focus on detecting differences between the two age groups regarding their mental representation of the cellular phone's menu and show the relationship between incorrect or incomplete mental representations and the performance actually using the device. Results always include users of both cellular phones models, since differences between the two phones are not of central interest here. First, the users' survey knowledge is analyzed, then the procedural knowledge, and finally, landmark knowledge.

3.1 Did Participants Have a Correct Representation of the Overall Structure of the Phone Menu? Survey Knowledge

Analysis of the card sorting task revealed that 4 of the older subjects did not arrange the cards in a hierarchical structure. Instead, one subject arranged the cards in clusters of three, without an interconnection between the clusters, possibly, because he simply mirrored the arrangement of menu functions he had seen on the display (always three menu items were presented at a time). Two participants had no idea at all of how to arrange the cards because they could not imagine what was meant with the functions on the cards or how a menu could be organized. One user explained that it would have been the easiest if each function were allocated to a specific key. In the younger group, on the other hand, all participants laid a hierarchical menu structure (Table 1).

Table 1. Number of users who laid a hierarchical and a non hierarchical menu structure in the card sorting task

	Mental representation of cellular phone menu	
	hierarchical	non hierarchical
20-32 years ($N = 16$)	16	0
50-64 years ($N = 16$)	12	4

To analyze the impact of having a mental representation of the menu's tree structure on the ability to effectively and efficiently interact with the device, the performance of those who laid a hierarchical structure in the card sorting task and those who did not is compared.

Taking only the older subjects, it was shown that the 12 users with a mental representation of the tree structure solved on average 80.2% of the tasks (3.2 out of 4 tasks) while the others solved only 65.6% (2.6 tasks). This difference yielded statistical significance ($t(14) = 2.43$; $p < .05$). When considering all 32 participants, the difference between the two groups was even somewhat bigger with users who had a correct mental representation solving on average 89.7% of the tasks (3.6 tasks), thus 24.1% more tasks than the users who were not aware of the tree-structure ($t(30) = 3.96$; $p < .001$) (see Figure 3, left). The awareness of the hierarchical structure of the menu also had an effects on the time users needed to process the task (Figure 3, right), though not yielding statistical significance.

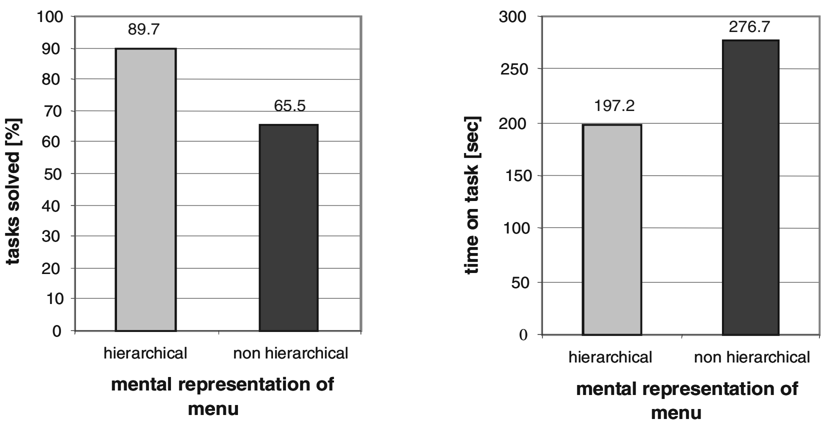


Fig. 3. Performance using the cellular phone depending on a hierarchical mental representation of its menu; N=32 participants (16 between 20 and 32, 16 between 50 and 64 years)

The correctness of the mental representation of the menu structure – the survey knowledge of the menu - can also be expressed in the number of levels the participants used to structure the cards. The branch of the menu, which had to be clustered, consisted in fact of four levels in both phones. Of the older participants, only one arranged the cards in four levels, while in the younger group 6 persons did so (Table 2).

Table 2. Number of users who laid 0 to 5 levels in the card sorting task

	Number of levels in the mental representation of the menu				
	0-1 level	2 levels	3 levels	4 levels	5 levels
20-32 years ($N = 16$)	0	3	5	6	2
50-64 years ($N = 16$)	4	9	1	1	1

Comparing the group with a correct mental representation of the depth of the menu with the rest of the participants regarding their performance solving the phone tasks, it is shown in Figure 4 that they not only solved more tasks (96.4% compared to 84%,

(30) = 2.24; $p < .05$), but also needed less time to process these tasks (109.9 sec compared to 234.4 sec, $t(30) = 2.95$; $p < .01$).

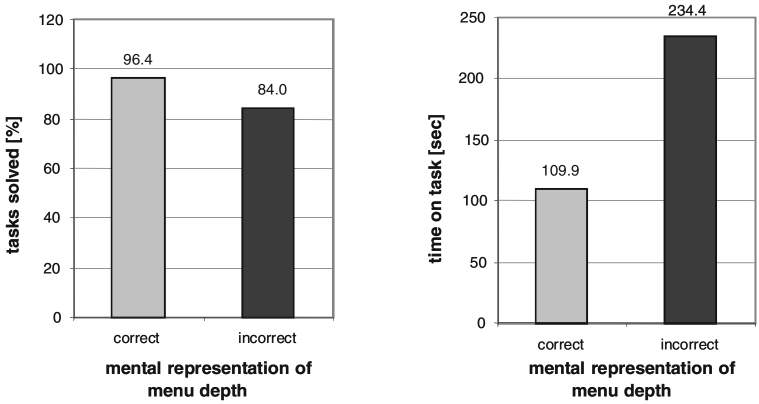


Fig. 4. Performance depending on the correct mental representation of the cellular phone menu's depth; N=32 participants (16 between 20 and 32, 16 between 50 and 64 years)

The one person in the older group with a correct mental map of the menu depth solved 100% of the tasks and needed 140.5 seconds for that, while the remaining older participants solved only 75% ($t(14) = 2.29$; $p < .05$) and needed double the time to process them (281.1 sec, $t(14) = 1.7$; $p = .1$). This person with a correct mental representation of the menu depth thus met the performance of the younger participants using the mobile phone who solved on average 96.9% of the tasks in 142.1 seconds.

Older participants arranged the cards in general in a much shallower structure (see Table 2), on average 2.1 levels, while younger subjects structured on average 3.4 levels ($t(30) = 3.6$; $p < .01$), thus being much closer to the correct depth of 4 levels. With a less strict criterion of the correct menu depth – 4 +/- 1 levels - users can be divided into two groups of equal size. 16 participants had a mental representation of the phone's menu consisting of 3, 4 or 5 levels, 16 subjects thought the menu consisted of only two levels or did not think of a hierarchical structure at all. Only three older participants had met this criterion of laying 4 +/- 1 levels. Again it could be shown that even with this less strict criterion meaningful performance differences can be found. Participants laying 3, 4 or 5 levels solved 94.5% of the tasks taking 167.4 seconds, while the participants who structured the cards in 2 or less levels solved only 78.9 % ($t(30) = 3.84$; $p < .01$) and took 246.9 seconds for the processing of the 4 phone tasks ($t(30) = 2.15$; $p < .05$). The task of hiding their own number when calling, which was the most difficult and the one relevant for the card sorting task, was only solved by 5 of the 16 participants who structured less than three levels, while of the subjects with a more accurate mental model of the menu depth, 13 persons solved the task ($t(30) = 3.20$; $p < .01$).

In summary, older participants showed to have an inferior survey knowledge of the phone menu than younger users, with not only a more shallow notion of the menu's

depth, but sometimes not even a hierarchical representation. It could be shown that this inferior mental model was indeed associated with a poorer navigation performance.

3.2 Is the Path to Be Taken Represented in User's Mental Model? Route Knowledge

The correct route from "Settings" to the point in the menu where the function of hiding the phone number has to be set "on" was only structured correctly by two participants, both belonging to the younger group (see Table 3). These two persons solved 100% of the tasks correctly while the rest solved on average only 85.8% ($t(29) = 5.61$; $p < .001$). When taking a less rigid criterion, namely that at least three of the four functions of the path to be structured are correct, still only 2 of the 12 persons who accomplished it are older subjects. Again, it could be shown that those with a nearly correct mental map of the route were able to solve more tasks correctly (95.8%) than the others (81.3%, $t(30) = 3.33$; $p < .01$) and needed significantly less time (139.8 sec compared to 247.6 sec, $t(30)=3.0$; $p < .01$) (Figure 3).

Table 3. Number of users who laid the correct route in the card sorting task

	Correct mappings in the mental representation of the route				
	0	1	2	3	4
20-32 years ($N = 16$)	0	1	5	8	2
50-64 years ($N = 16$)	6	4	4	2	0

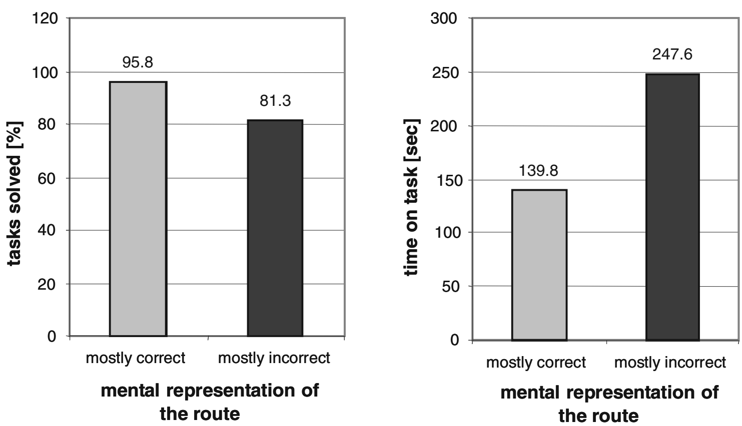


Fig. 5. Performance of users depending on their route knowledge; N = 32 participants (16 between 20 and 32, 16 between 50 and 64 years)

Restating, it was demonstrated that route knowledge of the path in the phone menu to be selected is distinctly less represented in older participants' mental model than in

the younger participants'. This may explain older users' inferior performance as it was shown that better representations of the route go along with the higher ability to solve tasks on the phone in a shorter period of time.

3.3 Are Salient Features of the Menu Branch Mentally Represented? Landmark Knowledge

Landmark knowledge is defined here as the total number of correct mappings in the card sorting task. That is, the number of functions that have been correctly allocated to the corresponding superordinate term or to the first menu level. Apart from the right route to be taken these landmarks are also important for orientation as they can further indicate which way *not* to take in order to solve a specific task. Regarding landmark knowledge, older adults once again turned out to be inferior to the younger adults. The older group allocated on average only 4.5 of the 22 functions correctly, while in the younger group 11.4 cards were arranged to the right position within the menu. This difference is highly significant ($t(30) = 5.8$; $p < .001$).

The importance of landmark knowledge, or which function is to be found under which category within the cellular phone menu, for successful interaction with the device was further demonstrated in the study: Correlations between the number of functions allocated to the right superordinate term and the percentage of tasks solved was $r = .78$ ($p < .001$), with time on task $r = -.65$ ($p < .001$). This means that the better the users' landmark knowledge of the menu structure, the more tasks they solved and the less time they needed (see Figure 6).

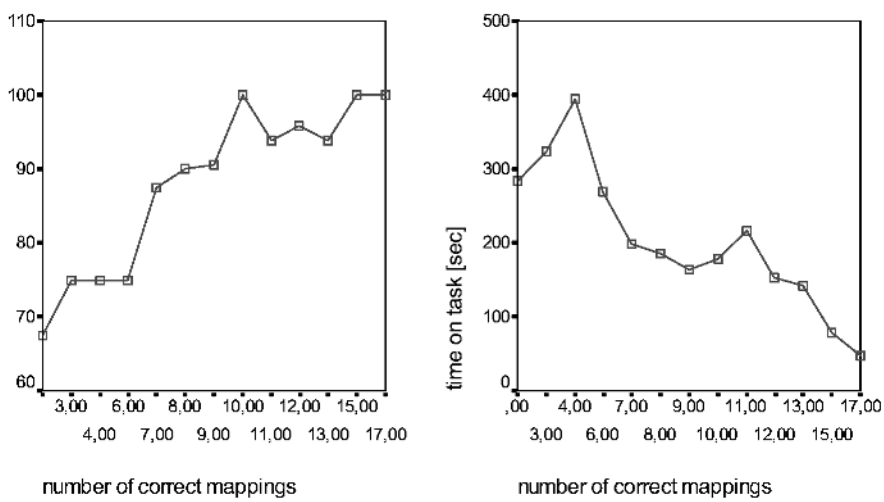


Fig. 6. Performance of users depending on their landmark knowledge; $N = 32$ participants (16 between 20 and 32, 16 between 50 and 64 years)

3.4 Comparison of Results with Children's Mental Models

In a similar study the same tasks were applied to children and teenagers aged 9 to 16 [2]. In contrast to the adults, however, the kids processed the four tasks on the cellular phones twice. The kids' mental representation of a cellular phone's menu is of special interest because it is assumed that due to their contact with technology from early on, they should have a fairly correct notion of its structure. Contrary to the expectation, 3 of the 21 participants did not cluster the functions in a hierarchical menu tree, thus resembling the *older* adults in their mental representation of the menu. The remaining 17 subjects however, had a fairly correct mental representation of the menu depth, with on average 3.7 levels. Accordingly, their survey knowledge was better than the young adults', but one has to take into consideration that the kids had processed the tasks twice before structuring the cards. The whole route was mapped by 7 participants correctly, thus again outperforming both adult groups, and landmark knowledge with on average 10.2 correctly allocated functions was nearly as good as in the young adult group.

As in the adult group, a meaningful superiority in the performance using the cellular phone of participants with a more accurate mental map of the menu was found. Children who were aware of the hierarchical nature of the menu solved 92.5% of the tasks compared to 75% in the remaining group. Children with a correct representation of the route needed 85% less time than the rest to process the relevant task of hiding their own number. Thus route knowledge showed to have the greatest influence on the users' performance using the phone. The correlation between landmark knowledge and time on task on the other hand, was smaller than in the present experiment, with $r = -.48$.

4 Discussion and Conclusion

In the present study it was demonstrated that users' mental model of how a mobile phone menu is structured significantly influenced their navigation performance. Crucial for the performance using a mobile phone is the knowledge that functions are arranged hierarchically (survey knowledge), the representation of the menu depth (route knowledge), as well as memorizing under which superordinate term each function is located (landmark knowledge). Thus, the three types of knowledge involved in spatial orientation in the natural environment [10] are also of importance for successful interaction with a cellular phone possessing a hierarchical menu structure. Further, it was corroborated that younger and older users' mental models differ substantially. Older adults' mental model of the menu was not always hierarchical, but instead linear or functions were arranged in clusters without any interconnection. Moreover, seniors showed to have a more shallow representation of the menu and allocated fewer functions correctly to superordinate terms. The specific attributes of older users' mental representation resulted in inferior navigation performance compared to the younger group.

The factors that determine older users' lower competency handling the cellular phone should be considered when designing cellular phones which are supposed to meet the demands of a broad user group: First, the declining memory capacity, which probably leads to difficulties learning "landmarks"; that is, the location of functions within the menu. Secondly, as spatial abilities decline over the life span, it is even more difficult for seniors to orient themselves in the menu-tree, especially when hidden from sight as it is the case in a cellular phone due to the small display. A third important aspect may be the fact that most older adults are less experienced with technology, such as menu-driven devices.

As a comparison with the results of a study analyzing children's mental model [2] has shown, however, the difficulty of building an appropriate map of the hierarchical menu structure is found not only in older people but also in the very young generation, which has grown up with technology from. The fact that we did not encounter this problem in our young adult group may be ascribed to having a sample of university students taking part in this study (and many other usability studies). Findings gathered with this user group should not be simply generalized for the broader population.

The findings presented here may have implications for the design of cellular phones in general. First, the inherent menu structure seems not to be transparent to older users – even if they are used to working with programs such as Windows Explorer, which is organized in the same fashion as our sample. One way of overcoming problems associated with hierarchical menu structures which was proposed in a recent study [8], is to use only one long alphabetical list of functions, where users can search by initial letters. This was evaluated with students. It is to be questioned whether this really helps users less experienced with mobiles as they often have no idea of the functions' naming in the menu and simple recognition of functions and categories – even though far from trivial – should be easier than active recall of the right term for a specific function.

It may be concluded therefore that the usability of a traditional mobile phone definitely has to be improved to meet demands not only of younger adults but of a broader user group, including children and seniors. One way of providing helpful navigation information is to make the menu structure more transparent through graphical hints on the display or in the manual. The positive impact of showing users a handout with the mobile phone's complete menu tree including the route to be taken to solve a task was already demonstrated [1]. But as most users prefer not to read the manual [6], this alternative will probably not have a strong impact on the usability of the device. The graphical hints on the display of phones currently found on the market differ distinctly regarding the type of information, such as the degree of survey knowledge, they provide. Headings, indicating which sub-menu was selected, provide some form of landmarks; scrollbars, for example, show where the user is currently located within a specific level of the menu, without providing neither landmark nor survey knowledge about the overall structure. Numbers displayed in a corner of the screen, representing the selected functions on each level (e.g. 3-2-4), provide information about the depth and breadth of the menu (survey knowledge) including the current location within the overall structure and how to get there (route knowledge),

but are rather abstract. It will be very instructive to find out which of these (or alternative) forms of visualization helps users most, possibly compensating for the ongoing decline of cognitive functions in seniors.

The findings of the present experiment also have implications for a support designed to help elderly users when starting to use a device like the cellular phone – which has been claimed by Tuomainen & Haapanen [11]. Here, explaining the hierarchical tree structure of the cellular phone's menu should prove to be very helpful.

The fact that, especially in older participants, landmark and route knowledge of the cellular phone's menu is rather poorly mentally represented may, as already mentioned, be ascribed to the large memory load imposed on the user by the current design realizations of the devices. Therefore it is especially important for this user group to have unambiguous naming and allocation of functions to submenus and categories in order to decrease memory load. This issue is currently under study.

Acknowledgements. We acknowledge the participation of René Müller who collected and analyzed the data as well as Hans-Jürgen Bay for helpful comments on earlier versions of this paper.

References

1. Bay, S.: Cellular phone manuals. Users profit from spatial maps. Proceedings of CHI, Conference on Human Factors in Computing Systems, Extended Abstracts (2003) 662-663
2. Bay, S., Ziefle, M.: Performance on mobile phones: Does it depend on proper cognitive mapping? In: D. Harris, V. Duffy, M. Smith and C. Stephanidis (Eds.). Human-Centred Computing. Cognitive, Social and Ergonomic Aspects. Vol. 3. Lawrence Erlbaum, Mahwah, NJ (2003) 170-174
3. Bay, S., Ziefle, M.: Design for all: User characteristics to be considered for the design of devices with hierarchical menu structures. In H. Luczak and K.J. Zink (Eds.) Human Factors in Organizational Design and Management. IEA Press, Santa Monica (2003) 503-508
4. Birren, J. E., Fisher, L. M.: Aging and speed of behaviour: Possible Consequences for psychological functioning. Annual Review of Psychology, 46 (1995) 329-353
5. Cerella, J. Poon, L. W., Fozard, J. L.: Mental rotation and age reconsidered. Journal of Gerontology, 36, 5 (1981) 620-624
6. Jordan, P.W.: An introduction to usability. Taylor & Francis, London (1998)
7. Maguire, M., Osman, Z.: Designing for older and inexperienced mobile phone users. In: C. Stephanidis (ed.). Universal Access in HCI: Inclusive design in the information society. Lawrence Erlbaum, Mahwah, NJ (2003) 439-443
8. Marsden, G., Jones, M.: Ubiquitous computing and cellular handset interfaces – are menus the best way forward? South African Computer Journal, 28 (2002) 67-76
9. Salthouse, T. A., Prill, K. A.: Effects of aging perceptual closure. American Journal of Psychology, 101 (2) (1988) 217-238
10. Thorndyke, P. W., Goldin, S. E.: Spatial learning and reasoning skill. In: H. L. Pick & L. P. Acredolo. Spatial Orientation. Theory, research and application. Plenum, New York (1983) 195-217

11. Tuomainen, K. & Haapanen, S.: Needs of the Active Elderly for Mobile Phones. In: C. Stephanidis (ed.). Universal Access in HCI: Inclusive design in the information society. Lawrence Erlbaum, Mahwah, NJ (2003) 494-498
12. Vicente, K. J.; Hayes, B. C.; Williges, R. C.: Assaying and isolating individual differences in searching a hierarchical file system. Human Factors, 29 (3) (1987) 349-359
13. Zaphiris, P.: Age differences and the depth-breadth tradeoff in hierarchical online information systems. In: C. Stephanidis (Ed.). Universal access in HCI: Towards an information society for all. Lawrence Erlbaum, Mahwah, NJ (2001) 540-544

Using Landmarks to Support Older People in Navigation

Joy Goodman, Phil Gray, Kartik Khammampad, and Stephen Brewster

Glasgow Interactive Systems Group, Department of Computing Science
University of Glasgow, Glasgow G12 8QQ
{joy, pdg, stephen}@dcs.gla.ac.uk

Abstract. Although landmarks are an integral aspect of navigation, they have rarely been used within electronic navigation aids. This paper describes the design of a pedestrian navigation aid for a handheld computer, which guides the user along a route using photographs of landmarks, together with audio and text instructions that reference these landmarks. This aid was designed with older users in mind who often find their mobility hampered by declines in sensory, cognitive and motor abilities. It was tested against the standard paper map for the test area with both younger and older people and their performance and subjective workload were measured. The results show that such an aid can significantly outperform a paper-based map and that older people derive substantially more benefit from it than do younger people.

1 Introduction

The proportion of older people in developed countries is rapidly increasing [13], producing an urgent need to provide greater support for this section of the population. This concern, together with the possibilities and challenges posed by this user group, has prompted recent research into ways in which technology can support and include older people. Much of this research has focused on indoor and stationary applications (see, for example, [14]), but recent advances in handheld computers and positioning technology mean that there is great potential to support older people in mobile situations as well.

Navigation is an important mobile activity, key for maintaining mobility and independence. However, many older people find increasing difficulties with it due to declines in their perceptual, cognitive and motor abilities [8]. This is therefore one area where technology could make a positive difference, through the use of computerised navigation guides and aids.

Current navigation guides usually support navigation by guiding the user along a given route, using turn-by-turn arrow-based directions, or by presenting maps (see, for example, [4]). Some research projects have tried other methods, such as overlaying information on a detailed 1st-person view of an area (e.g., [9]), thus allowing reference to specific environmental information, such as landmarks.

However, landmarks have rarely been used explicitly or in a consistent fashion in navigation guides, although landmarks are an integral aspect of navigation itself [11].

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 38–48, 2004.

Burnett has shown that their use in vehicle navigation systems can greatly improve their effectiveness [1], but their use within pedestrian navigation guides has not been fully investigated.

This paper describes the design of a pedestrian navigation aid that uses landmarks to help guide the user along a route. This device was designed with older users in mind and used to investigate the feasibility of using landmarks in such a device and the possibilities for how these can be presented. Its performance was compared with that of a paper-based map and age differences in its use were analysed.

2 The Design of the Navigation Aid

The design of the navigation aid was informed by some preliminary requirements gathering in the form of focus groups with older people [5]. They answered questions and completed exercises on their travel, navigation and ways of giving directions. Among other results, we found that participants appreciated information about landmarks and liked to be given a visual indication of what these landmarks looked like.

We therefore designed a prototype aid that describes routes using photographs of landmarks. The landmarks were also presented using text and audio directions, so that participants' reactions to the use of landmarks would not be determined by any one particular method of presentation and so that responses to different methods could be gauged.

Ultimately, such an aid would be incorporated into a larger system, allowing users to explore an area and select different start and end points, adapting routes to different users' requirements and coping gracefully when users wander away from the route. It could also be included as part of a larger system that provides, for example, information about the surroundings and places of interest. However, this study focused on the core aspect of such a device (the navigation aid itself) and the methods by which navigation assistance can be best provided.

2.1 Implementation and Interface Design

The application was written in C# and deployed on a Compaq iPAQ. A sample screen is shown on the left in Figure 1. It displays a photograph (58x43mm) of a landmark that can be seen from the start of the route. Once the user reaches that location, he or she should press the button labeled "Next Image" to progress to the next instruction and receive a photograph of a new landmark or location to head towards.

As well as the photograph, a brief text instruction is shown and a longer speech instruction can be heard when the "Audio" button is pressed. For example, the speech instruction for the screen in Figure 1 is "From the Western Lecture Theatre, if you look right, you'll see a large chimney. Please go up to it." Audio is presented using the device's built-in speaker. The "View Map" button shows a simplified map of the route as shown on the right in Figure 1. The position of the landmark in the photograph is marked with a red dot (dark grey and circled in Figure 1). The "Restart" button returns the user to first screen at the beginning of the route.

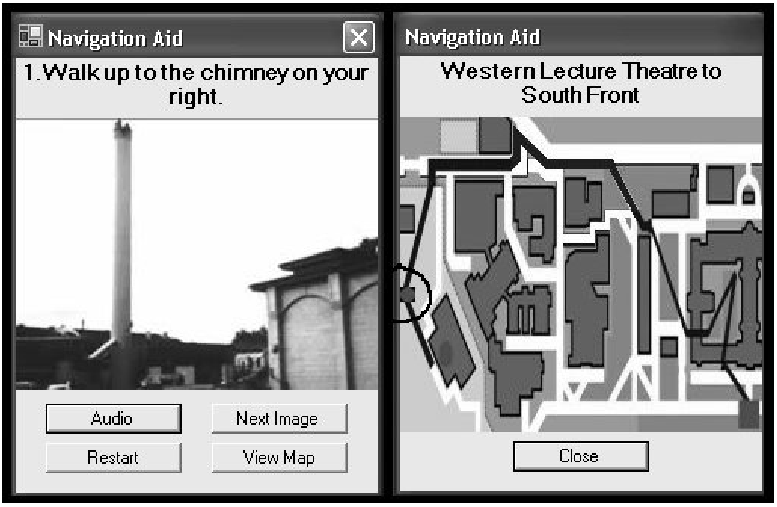


Fig. 1. Example screens from the navigation aid

The interface was designed with guidelines for the design of desktop applications for older adults in mind (e.g., [7]), as little work on mobile design guidelines for older people had been carried out. Although these guidelines had to be adapted to take account of the limited screen size and different interaction techniques, they provided a useful starting point for the design.

For example, drop-down menus and unnecessary features were avoided, text rather than icons was used on buttons and sans-serif fonts were used. A slightly smaller font size (11pt instead of 12pt) than that recommended in [7] had to be used due to the limited screen size available. To compensate for this, a **bold** typeface was used. In addition, a natural male voice with a standard accent was used for the audio instructions, in line with findings from Lines and Hone [10].

The interface was kept particularly simple for the purposes of testing. In a fully functional system, more attention could be paid to enhancing the aesthetic appeal of the device. However, care must be taken not to compromise the usability of the device, as older people are more easily distracted by un-necessary screen elements and irrelevant information [15].

3 Evaluation Design

3.1 Field Experiments

The navigation aid was tested using a set of field experiments. Field studies were necessary because the navigation aid is highly dependent upon the surrounding environment and cannot be tested realistically in a laboratory setting. An experimental setup was used to obtain a quantitative comparison of our device against a paper map and of usage by different age groups.

We choose to use an experimental evaluation rather than an ethnographic field study in order to focus on a single aspect of the device (the aiding of navigation) and to obtain quantitative comparisons of performance between different groups of users.

However, field experiments do present the experimenter with several challenges, primarily that of limiting the effect of possibly confounding variables, such as light and noise levels, weather conditions and the time of day. When the levels of such variables cannot be kept consistent, their effects on results can be reduced by varying them across conditions when possible. Removing all variation, however, would produce unrealistic results which may not mean anything for real-world usage. Using real locations and realistic environmental conditions gives real data on how the device is used in practice, and the advantages outweigh the difficulties and make the extra effort worthwhile.

3.2 Participants

The navigation aid was tested with 32 able-bodied users; 16 aged between 63 and 77 and 16 between 19 and 34. Each group was balanced with respect to gender. In addition, four "backup" participants were recruited and run through the experiment. This was done in case any data from the main participants was confounded by large changes in the external environment. Only data from one backup participant was actually used because one of the main participants arrived late and so did the experiment in the dark.

In order to avoid over-familiarity with the area, part of the campus of Glasgow University, no participant was either a student or staff member at that university.

All but one of the older participants had never used a hand-held computer before and the remaining participant had only used one a few times. The majority of younger participants (11 out of 16) had also never used a hand-held computer and only one was a regular user. All participants had used a map before with 10 of the younger participants and 11 of the older participants rating themselves as regular map users.

3.3 Method

Participants were asked to navigate along two different routes, one of them using the device and the other using the standard paper map for the area. The order of the two routes and the two methods (device or map) were counterbalanced, creating four conditions. Equal numbers from each age group and gender were assigned to each condition.

The experiment was conducted on part of the campus of the University of Glasgow, a common tourist destination within the city of Glasgow. This location was chosen because it has a large number of junctions and decision points in a small area, allowing a sufficiently complicated route to be tested while limiting the length of the routes (and therefore of the experiment) in order to avoid tiring participants, particularly the older ones. It also has a low volume of traffic, creating a relatively safe environment, and thus conforming to ethical guidelines.

Routes were chosen within this area with 13-16 waypoints and taking about 10 minutes to walk if walked directly. The sequence of photographs in Figure 2 illustrates a segment of one of the routes used.

Fig. 2. Images from one of the test routes

3.4 Map Condition

The map used was a greyscale version of that available at http://www.gla.ac.uk/general/maps/colourmap.pdf, covering a slightly smaller area and with a shorter list of buildings. A route was indicated on the map using a sequence of numbered, highlighted circles. Part of this map is shown in Figure 3.

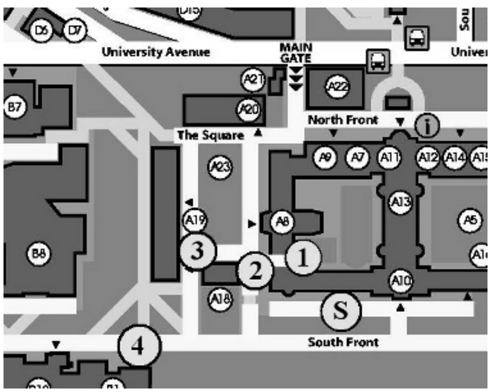


Fig. 3. Part of the map used in the experiment (close to real size)

Participants were asked to navigate along the indicated route, visiting each of the numbered locations in turn. The equivalent route when using the device also passed through each of these locations in the same order.

3.5 Procedure

After an initial briefing, the use of the map or device was explained to participants and they then used this method to find their way along the route. Each participant navigated both routes, using the map on one and the device on the other.

On the routes, the experimenter walked with the participants, a few steps behind them in order not to influence navigation. He or she made written observations on navigation behaviour, as well as providing help when participants got lost. Such help was only provided when it was necessary. Help was given to prevent distress and conform to ethical guidelines, and was noted by the experimenter.

After each route, participants filled in a questionnaire on the device or map. This incorporated the NASA Task Load Analysis (TLX) scale [6], which measures perceived workload, as an indication of how the participants felt about using these methods. We modified the TLX response scales slightly to provide only five possible responses to make them simpler and less daunting for older participants.

The time taken and the number of times that participants got lost on each route were also measured. A participant was defined as lost if the experimenter had to intervene.

4 Evaluation Results

4.1 Timings and Frequency of Getting Lost

The mean times taken to navigate the routes with the map and the device are shown in Figure 4. A two-way ANOVA on age and method showed a significant main effect of both the navigation method (map or device) and the age group and, perhaps most illuminating, a significant interaction between age and method (all $p<0.001$).

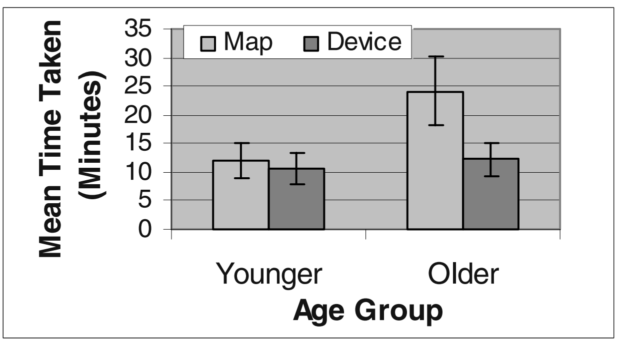


Fig. 4. Mean time taken to navigate test routes (error bars show standard deviation)

The navigation method affects the time taken; participants took significantly less time when they used the navigation aid than when they used the map. Age group also affects the time, with younger participants being significantly faster.

However, both of these effects do not really describe the results shown in Figure 4. The interaction between the different factors gives a much clearer insight into the results. An analysis of the interaction showed that the time differences between the age groups are only significant with the map, not with the device, and that only the

older sample displays a significant difference between the map and the device ($p<0.001$, t-tests).

Participants also got lost significantly less often with the device ($p<0.001$, t-test[1]), where "lost" is as defined above. In fact, no participants got lost when using the device, compared to a mean of 1.9 times per route for older users and 0.4 times for younger users when using the map.

4.2 TLX Scores

Raw TLX (RTLX) scores were calculated as a measure of overall workload [2]. These were significantly lower for the device than for the map ($p<0.001$, Mann-Whitney) and there was no significant effect of age group ($p>0.05$). The TLX scores can be further investigated by analysing their individual components as shown in Figure 5. Because we simplified the TLX scales, these scores were calculated out of 5 instead of 20.

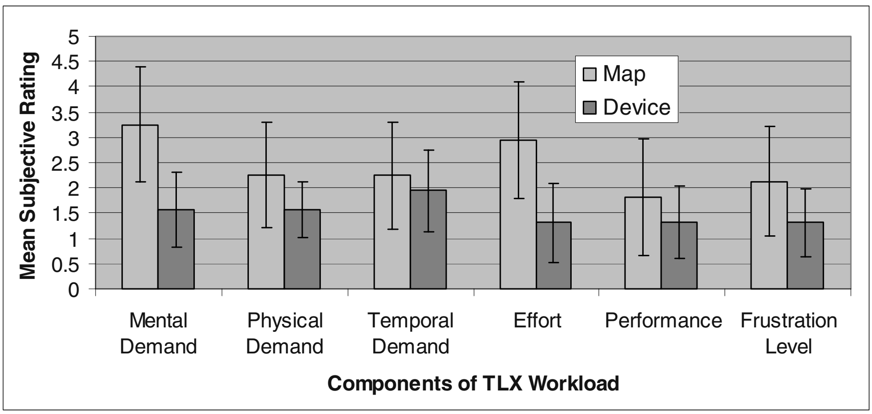


Fig. 5. Mean TLX Scores for the map and navigation aid. Higher values indicate higher workload and lower performance. Error bars show standard deviation

Using Mann-Whitney tests, significant differences between the map and the device were found for mental demand, effort and frustration ($p<0.001$), as well as for performance ($p<0.005$) and physical demand ($p<0.05$). There was no significant difference in temporal demand ($p>0.05$).

[1] Although the frequency of getting lost is non-parametric, a t-test was used because tests such as Mann-Whitney could not be applied since all of the results for one of the conditions (use of the device) are identical (all are 0). In this case, the results from the t-test are significant enough and a t-test robust enough to conclude that a significant difference does exist.

4.3 Preferences and Comments

After trying both methods, participants were asked to indicate on a 5-point scale which method they found most useful. The results are shown in Figure 6 below. Only one person (an older user) indicated a preference for the map, explaining that she was "accustomed to using maps and feels comfortable with them".

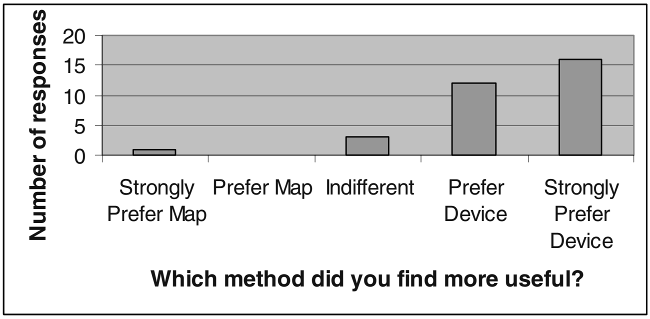


Fig. 6. Perceived relative usefulness of the map and the navigation aid

Reasons given for preferring the device were varied, with some people giving multiple reasons. Most commonly mentioned was the provision of images of locations, which some said helped them to confirm where they were or to determine more easily where they should go. Some participants liked having step-by-step directions and one user said this was like "walking with a guide who knew each and every corner".

Other reasons focused on the shortcomings of maps, with both difficulties with maps in general and specific shortcomings of the map provided being mentioned. One participant explained that "using maps in general is quite difficult", while another described how he "would have to turn and try and work out which way using the map". A few people mentioned shortcomings of the particular map used; for example, one person complained that there was no indication that a grey area represented a car park.

Despite the preferences for the device, some of the participants did have some reservations about it. Some felt that a map would be better for longer routes; others that the device gives less freedom and control over the route and a poorer idea of the route as a whole. Nevertheless, the majority felt the device to be more useful.

5 Discussion

This work demonstrates that landmarks can be used effectively to support navigation through a handheld device. Such a device can improve the time taken to navigate a route and reduce the number of times when people get lost, compared to a paper map. It can also reduce workload and users agree that the device is more useful than the map.

Furthermore, such a device has a greater potential benefit for older users than for younger. Although both older and younger users found the device useful (they got lost less often with the device, found that it produced a smaller workload and felt that it was more useful than the map), only older participants completed routes faster when using the device.

It may have been expected that older participants would have difficulty using a handheld device, particularly because all but one had never used one before. However, participants had little difficulty using the navigation aid and gave it low ratings on all aspects of TLX workload. We found that if the interface on a handheld computer is carefully designed with older people in mind, then it can be used without difficulty by this age group. This agrees with results from other studies, such as [12]. In [12], McGee et al found that, when the interface for a handheld application for cancer patients was redesigned based on the results of pilot studies, it was used without difficulty by the user group, many of whom were in the older age category.

We also expected older participants to be slower than younger ones, due to reduced walking speed, but this effect was *only* observed with the map. A landmark-based device has the potential to improve older people's performance to a level comparable with a younger age group.

This does not mean that such a device would not be useful for any younger person. Although the younger group as a whole did not experience significant improvement with the device, individuals did. We cite the example of one of the backup participants (one of the younger group) who took over three times as long and got lost 7 times with the map, as opposed to once at the start when using the device.

The use of the device can be investigated further by examining the components of the TLX scores. As well as increasing efficiency, the device has the advantage of decreasing the mental and physical demand, the effort expended and the frustration experienced. Users were also aware of an increase in their performance level.

There are a variety of possible reasons for this increase in performance and decrease in workload. While this study cannot give any definitive answers, some indications can be gathered from participants' comments. When asked to explain why they preferred the device, several participants (both older and younger) explained that it gave a visual identification or confirmation of locations on the route. Several also liked being given a set of directions and being told which direction to turn in rather than having to figure it out from a map.

All of this does not mean that there are no difficulties with a landmark-based navigation aid. The step-by-step nature of such an aid reduces the user's freedom and control and provides a poorer overall idea of the route. There is also a degree of natural resistance to new methods. Research is needed into ways to overcome these challenges, e.g., by providing support for the user to change the route.

6 Future Work

Although we have demonstrated that landmarks can be used effectively in our test area, this area is only representative of a subset of possible locations. Landmark-

based navigation aids also need to be tested in environments such as city centres and shopping areas, which have different kinds of street layouts and in which people may be doing different kinds of activities. Similarly, the map used in this study was the standard map provided by the university to its visitors. While similar in type to a street map, there were some significant differences, as it used a simplified and stylised design, rather than being topologically correct. It is important to see how navigation aids compare with standard street maps.

It is also unclear how much of the success of the navigation aid can be ascribed to different factors, such as the electronic nature of the aid, the step-by-step directions, the images of landmarks and the verbal and written instructions. In particular, there are many ways of presenting landmarks to users. The aid used in this study did this through photographs, text and speech in an electronic medium, but other methods are possible. It is important to examine the different methods and modalities to determine which are most effective in aiding navigation.

As a first step towards this, we have modified the navigation aid from this study to provide a choice of different methods of presenting the information: as well as a combination of photographs, text and speech, there is a text only, a speech only and a text and speech interface. We are currently conducting a set of field experiments, comparing these last three versions of the interface, in order to analyse the contribution of each modality.

We also hope to investigate ways in which a navigation aid can help users to gain a better overall idea of a route and area, perhaps enabling them to explore a location rather than follow specific routes. In the course of this, we plan to consider issues of users' freedom and control over their navigation and the routes that they take.

7 Conclusions

Landmarks are a key part of navigation and this study has shown that they can be used effectively within electronic pedestrian navigation aids. A device that bases its navigation guidance around landmarks can significantly outperform a paper-based map, as well as reducing subjective workload and eliciting a positive response from users.

In addition, we have found that older people derive substantially more benefit from such a device than do younger users, with a large reduction in the time taken to navigate routes. The use of handheld technology does not prevent them from using the navigation aid successfully. Such aids could, therefore, provide key support to older people in maintaining their mobility and independence.

Acknowledgements. This work was funded by SHEFC through the UTOPIA project (grant number: HR01002), investigating the design of usable technology for older people [3]. We would like to thank Morag Stark for helping to conduct some of the experiments and also all those who took part in the experiments.

References

1. Burnett, G.E.: "Turn right at the traffic lights" The requirement for landmarks in vehicle navigation systems. The Journal of Navigation 53(3) (2000) 499-510
2. Byers, J.C., Bittner, Jr, A.C. and Hill, S.G.: Traditional and Raw Task Load Index (TLX) Correlations: Are Paired Comparisons Necessary? In: Mital, A. (ed.): Advances in Industrial Ergonomics and Safety I, Taylor and Francis (1989) 481-485
3. Eisma, R., Dickinson, A., Goodman, J., Syme, A., Tiwari, L., Newell, A.F.: Early User Involvement in the Development of Information Technology-Related Products for Older People. To appear in Design Principles to Support Older Adults, Special Issue of Universal Access in the Information Society 3(2), to appear 2004
4. Garmin. http://www.garmin.com
5. Goodman, J., Dickinson, A., Syme, A.: Gathering Requirements for Mobile Devices using Focus Groups with Older People. To appear in Designing a More Inclusive World, Proceedings of the 2nd Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT), Springer (2004)
6. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P., Meshkati, N. (eds.): Human Mental Workload. North Holland B.V., Amsterdam (1988) 139-183
7. Hawthorn, D.: Possible implications of aging for interface designers. Interacting with Computers 12(5) (2000) 507-528
8. Kirasic, K.C.: Ageing and spatial behaviour in the elderly adult. In: Kitchin, R., Freundschuh, S. (eds.): Cognitive Mapping: Past, present and future. Routledge, Routledge Frontiers of Cognitive Science 4 (2000) Chapter 10
9. Laakso, K., Gjesdal, O., Sulebak, J.R.: Tourist information and navigation support by using 3D maps displayed on mobile devices. Proceedings of HCI in Mobile Guides, Workshop at Mobile HCI 2003 (2003) 34-39
10. Lines, L., Hone, K.S.: Older Adults' Evaluations of Speech Output, Proceedings of Assets 2002, ACM Press (2002) 170-177
11. Lynch, K.: The Image of the City. MIT Press, Cambridge, Mass. (1960)
12. McGee, M., Gray, P., Muir, L, Hargan, I: Designing a Chemotherapy Symptom Management System for PDAs: Emerging Design Guidelines from Early Patient Pilot Studies. Technical report TR-2003-146, Department of Computing Science, University of Glasgow (2003)
13. U.S. Census Bureau, International Data Base (IDB). http://www.census.gov/ipc/www/idbnew.html, updated 17th July 2003
14. Zajicek, M., Brewster, S. (eds.): Design Principles to Support Older Adults, Special Issue of Universal Access in the Information Society 3(2), to appear April 2004
15. Zec, R.F.: The Neuropsychology of Aging. Experimental Gerontology 30(3-4) (1995) 431-442

Heuristic Evaluation and Mobile Usability: Bridging the Realism Gap

Shirlina Po[1], Steve Howard[1], Frank Vetere[1], and Mikael B. Skov[2]

[1] Interaction Design Group, Department of Information Systems, The University of Melbourne. Australia
{showard,fv}@unimelb.edu.au

[2] Department of Computer Science, Aalborg University, Denmark
dubois@cs.auc.dk

Abstract. Heuristic evaluation (HE) is problematic when applied to mobile technologies, in that contextual influences over use are poorly represented. Here we propose two lightweight variants of HE: the Heuristic Walkthrough (HW) combines HE with scenarios of use, and the Contextual Walkthrough (CW) involves conducting the HW in the field. 11 usability experts were asked to use one of these three approaches to evaluate a mobile device and the usability flaws discovered were compared across technique. HW discovered more critical usability flaws than HE. CW revealed some unique problems relating to I/O and ambient lighting not encountered in the other two approaches. Though contextualizing heuristic evaluation improves the assessment of mobile devices, it appears that it is possible to introduce contextual detail, i.e. to bridge the 're-alism gap', with scenarios rather than expensive in-situ testing.

1 Introduction

Evaluating the usability of mobile devices remains a major concern in both research studies and industrial development projects. Mobile use strains our traditional, now well established, evaluation methods and tools. The consequences of decoupling interactive systems evaluation from the relevant context of use has been widely discussed and examined, but as we will argue, stripping a mobile device of its use context is especially problematic. We explore this general issue in relation to a provocative test case; a discount evaluation method often seen as acontextual (i.e. heuristic evaluation) and a situation rich in contextual challenge (i.e. mobile usability).

Our aim is to explore how contextual cues influence the conduct of a heuristic evaluation, and the consequent insights produced.

2 Related Work

Heuristic evaluation belongs to a category of usability evaluation methods which, together with the cognitive walkthrough, rely on guided expert assessment of the

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 49–60, 2004.

interface [11]. In heuristic evaluation the expert is guided by rules of thumb, and in the cognitive walkthrough a checklist and scenarios of use.

Along with scenarios of use and simplified think aloud techniques, heuristic evaluation is a central tool in so called lightweight, discount, or guerrilla approaches to usability engineering [14]. Such approaches do not stress cost minimisation alone, rather discount approaches are intended to be quick to learn, flexible in their application, and structured in such a way that the insights flowing from their application are continuos, rather than discrete and realised only at the completion of the usability cycle.

Heuristic evaluation is popular primarily due to this ease of implementation and efficiency [8]. Heuristic evaluation can be used early in the design cycle, and it is inexpensive, requiring neither end-users nor a functioning prototype [11]. Skilled usability evaluators can produce high quality results in a limited time. 3-5 experts have been known to identify 75-80% of known usability flaws [11]. For such reasons it is likely to continue to be a popular technique, especially in industrial applications, but will require ongoing refinement and extension.

The quality of the heuristic evaluation is highly dependent on the skills and experience of the usability experts. Heuristics are *"motherhood statements that serve only to guide the inspection rather than prescribe it"* [4]. Usability experts must exercise great judgment and interpretation skills in identifying issues. Further, heuristics are relatively product-oriented, they assess *"the system as a relatively self-contained object without strong contextualization in conditions of use"* [9]. Though the discount movement recognises the importance of scenarios of use in providing information on users, tasks and use contexts, many applications of heuristic evaluation take place in-absentia of such contextual detail. Simple minded application of heuristic evaluation is problematic as heuristics lack an awareness of the rather complex situational influences of the context.

What challenges do we face in evaluating use that is on-the run and that may last only a few seconds and is highly context dependent? Johnson [5] states that traditional approaches to evaluating stationary use may be inadequate as they reflect fixed contexts of use, single domains, with the user always using the same computer to undertake tasks. Dunlop and Brewster [3] contend that the mobile usability challenge relates to design for (i) mobility, (ii) a widespread user population, (iii) limited input/output facilities, (iv) incomplete and varying context, and (v) multitasking. These challenges are not simply those of interaction design; with the partial exception of (ii) and (iii), they are due to the vital and complex interplay between the real and the virtual and the need for our methods to consider the *"situation of use which is a consequence of the user being free to roam and having demand-access to resources"* [13].

A major methodological debate in mobile usability evaluation addresses the question of testing in the laboratory or testing in the field. Abowd and Mynatt [1] argue for testing in the field stating that 'deeper' evaluations require deployment into authentic settings. Furthermore, it is claimed that field studies provide an insight into aspects of the actual usage which are crucial for successful design of mobile technologies. However, Kjeldskov and Graham [6] found a strong bias in the current

mobile usability research for rigor over relevance with most papers using laboratory settings and a few papers using field settings or surveys. Instead of focusing on the contextual issues impinging on use, mobile evaluations tend to concentrate on device functionality [6]. Kjeldskov and Stage [2] confirm this finding and reports that between 1996 and 2002, only little mobile usability research included issues on usability testing. Their study found that less than half of 114 research papers on mobile usability included usability evaluations and these evaluations typically employed traditional techniques e.g. the think-aloud protocol collected in artificial laboratory settings lacking contextual cues and influence.

In evaluating mobile systems in artificial laboratory settings, activities in the user's physical and social surroundings can be difficult to simulate. Some studies have aimed to overcome this realism gap by recreating or simulating use context in the laboratory [6]. Kjeldskov and Skov [7] explored the influences of varying evaluation setting in a laboratory evaluation of a mobile collaborative device, and found that recreation of the use context had a significant impact on the results produced but that evaluation in-situ was problematic in terms of data collection processes, instrumentation and control of extraneous influences. Although aspects of use such as interruptions, complex patterns of cooperation and the physical and social environment are hard to simulate [10], they can also be difficult to observe; they may be particularly sensitive to the Hawthorne effect, or occur at times (e.g. night-time) or in places (e.g. private spaces) that are difficult to access.

Advances in ethnographically influenced usability approaches, even lightweight variants like Millen's rapid ethnography [15], are still predominantly used during the earlier user needs related phases of development and require substantial expertise in their application. Evaluation approaches that are sensitive to the contextual influences over use, and are inexpensive and flexible, would find ready customers amongst user experience professionals. It is reasonable to ask therefore, given the above, whether the benefits of in-situ evaluation outweigh the costs, and whether the contextual realism so desired by some authors, cannot be introduced by other means?

3 Approach

In order to further explore the interrelations between mobile use, heuristic evaluation, and the use context we conducted an empirical study that compared two ways of introducing contextual information into heuristic evaluation, against a baseline evaluation as described by Nielsen [11].

The baseline approach followed Nielsen's original ten heuristics, reproduced below.

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention

6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation

In addition to this baseline condition, we introduced contextual information in two simple ways (see [12] for details), producing three conditions:

- *The Heuristic Evaluation (HE).* A standard heuristic evaluation conducted in the laboratory. This setup carried no obvious contextual cues.

- *The Heuristic Walkthrough (HW)* combines heuristic evaluation with scenarios of use and the walkthrough is conducted in the laboratory. Thus the scenarios carried the contextual cues.

- *The Contextual Walkthrough (CW)* involves conducting the heuristic walkthrough in the intended situation of use. Thus both the scenarios and the situation that impinges upon the inspection carried contextual cues.

11 evaluators each used one of the three approaches - 4 used the HE, 4 the HW and 3 the CW (see Figures 1 and 2 for general setup). The assignment of experts to conditions took account of expertise in HCI, and familiarity with both heuristic evaluation and mobile devices. All evaluators had at least a semester of graduate level education in HCI, a number had doctorates in HCI, and all received further instruction in the techniques prior to data collection.

Fig. 1. Illustrative setup for HE and HW in the lab

Fig. 2. Illustrative setup for CW, scenarios 1 and 4 in the café

In the HW and CW conditions, the usability evaluators were given five scenarios to guide the assessment of the mobile device. These five scenarios were typical, realistic and covered the main functions of the Casio Cassiopeia™ E-10 pocket PC and consisted of (i) Creating a new appointment (see figure 3), (ii) Adding a new contact, (iii)

Creating a new task, (iv) Scheduling appointment within the week, and (v) Creating an alarm note.

For the CW evaluations, the locations were selected as realistic in terms of the scenarios, busy places around the campus of a major University, places filled with people, movement and noise. Scenarios one and five were conducted in a cafeteria, scenario two in a pub, three in an elevator, and four in an office.

Creating a New Appointment

Michael is a lecturer at the University of Melbourne. He teaches Financial Accounting and he has to record a lot of appointments and meetings to attend throughout the semester. As a result, Michael has bought a new Pocket PC in order to better organize his busy schedule.

Michael is currently in the canteen having lunch with his colleagues. Suddenly, a student approaches him and says that she has problems understanding the Accounting lectures. Since he is having lunch, he decides to have a discussion with the student another time. Therefore, Michael quickly reaches for his new Pocket PC and grabs the pointer (provided with the system) to key in data. He selects the appointment menu and enters the data and sets an alarm to remind him of the appointment when it is due. Once he finishes entering the details into the Pocket PC, he continues chatting with his colleagues.

Fig. 3. Sample scenario on creating a new appointment

All three conditions followed the same general protocol:

Stage One: Pre-evaluation Session

After greeting each evaluator, the goals of the study, the testing procedures, and the confidentiality issues were explained in detail. Scripts were prepared in advance and used for each usability evaluator to ensure consistency across experts and conditions.

In a demographics questionnaire experts were asked about their level of education, relevant experience in the field of HCI, experience in using both a PDA and Nielsen's heuristics. A training session was conducted with each evaluator to ensure that they fully understood the usability heuristics; this involved the facilitator stepping through each of ten usability heuristics and evaluators were invited to ask questions in order to clarify the meaning of each heuristic and their understanding of the overall process.

Stage Two: Evaluation

The usability evaluators performed the usability evaluation on the mobile device by identifying usability problems and prioritising them according to Nielsen's [11] five point ranking scales. A description of the ranking is given in Table 1.

While evaluating the mobile device, each usability evaluator was asked to 'think aloud' to explain what he/she was trying to do and to describe why he/she was taking the action. Their comments were audio taped and/or captured by one of the researchers.

The mobile device that was used for the evaluation in this study was the handheld Casio Cassiopeia E-10 pocket PC. A Personal Digital Assistant (PDA) was selected as a familiar and popular device for supporting multiple tasks, including personal information management. The Casio Cassiopeia is a handheld PDA running Windows

CE. The main features of the device include scheduling appointments, adding contacts and tasks, taking notes and so on. The user can either input characters by a pen using a QWERTY keyboard, or handwrite in a free form.

Table 1. Severity ranking scale

Rating	Description
0	I don't agree that this is a usability problem at all
1	Cosmetic problem only. Need not be fixed unless extra time is available on project
2	Minor usability problem. Fixing this should be given low priority
3	Major usability problem. Important to fix, so should be given high priority
4	Usability catastrophes. Imperative to fix this before product can be released

Stage Three: Debriefing
A short debrief discussion was held after each session, focused on the evaluators' experiences of the process, and providing an opportunity to probe where behaviour was implicit or puzzling to the researchers.

4 Results

Table 2 summarizes the quantitative data on flaws per evaluator, across each of the three approaches. HE identified 7.5 usability problems on average, HW 19 and CW 18.6. Note however that the variation within condition is significant, and suggests that the choice of evaluator is at least as important as the choice of technique. It should also be stressed that in tables 2 and 3, a strict comparison between conditions is difficult as the expression of what constituted a problem, especially its scope and granularity, varied between evaluators. We consider this issue further in Table 4.

Table 2. Number of discovered usability problems

	HE (N=4)	**HW (N=4)**	**CW (N=3)**
Participant	3	5	26
Participant	11	19	16
Participant	9	19	14
Participant	7	33	-
Total # flaws	30	76	56
Mean # flaws (SD)	7.5 (3.4)	19 (11.4)	18.6 (6.4)

On face value, situating the evaluation using scenarios increased the number of flaws caught, though there appears to be no additional benefit to immersing the evaluator in the context of use; however some indication exists that immersion reduced variance amongst evaluators.

Table 3 outlines the distribution of severity rankings across conditions. Consistent with the literature, HE discovered a large proportion of minor usability problems. HW identified more major problems than the HE, and HE failed to identify any usability catastrophes (as defined by the experts conducting the evaluations). This trend in the data suggests that the tendency of HE to focus on trivia can, in a limited way, be compensated for by using scenarios, though there is little to suggest that being in-situ adds further value. As compared with the HW, the CW identified fewer cosmetic problems.

Table 3. Distribution of severity ranking

	HE (N=4)	HW (N=4)	CW (N=3)
Cosmetic	21%	24%	9%
Minor	68%	34%	43%
Major	11%	33%	37%
Catastrophes	0%	9%	11%

Unsurprisingly there was variation between evaluators, even within condition, as to what constituted a catastrophic flaw. Therefore, a set of benchmark flaws was established by collapsing all reported problems across both experts and conditions, one of the authors then reclassified them according to severity. In this way we were able to identify five benchmark problems and then check whether the three evaluation approaches captured them. Benchmark flaws included both input (e.g. problems with keyboard), and output (e.g. 'reminder' option is hidden and obscured by visual keyboard) issues.

Table 4. Coverage of benchmark severe problems

	HE (N=4)	HW (N=4)	CW (N=3)
Benchmark Flaw 1	0	1	3
Benchmark Flaw 2	0	4	1
Benchmark Flaw 3	1	2	3
Benchmark Flaw 4	0	3	2
Benchmark Flaw 5	0	3	2

Table 4 outlines the frequency of benchmark flaws identified in each condition. The data suggests that, not only did HE discover fewer usability problems than the two adaptations; it also appears not to have caught the serious flaws (this is consistent

with the data derived from the 11 experts and presented in table 3). All three CW evaluators discovered the first and third flaws whereas all four HW evaluators discovered the second flaw. Furthermore, at least one HW and CW evaluator discovered all benchmark flaws.

So far we have dealt with the gross differences between the conditions. Now we will draw out some subtleties, especially as they relate to the types of problems revealed in each condition. Table 5 summarises this data. Each cell shows the number of flaws each expert identified within each heuristic[1]. The heuristics have been reordered to highlight those cases where a greater or lesser number of problems were reported for a given heuristic.

Table 5. Distribution of usability flaws across the 10 heuristics for each of 11 experts

		Heuristics 1-10									
		1	**2**	**3**	**10**	**7**	**8**	**4**	**5**	**6**	**9**
HE Experts	**1**	1	1	1	1						
	2	2	2	1		2		2			
	3	5	2	1		1			1		
	4	1	3	1	1		1				
HW Experts	**5**	1	1		1		2				
	6	3	5	5	2	1			3	1	
	7	3	4	1	1	2	2	2	1	3	
	8	7	4	3	1	4	3	3	2	4	1
CW Experts	**9**	5	7	3	1	1	3	2	2	1	
	10	2	5	1	2	1	2	3		1	
	11	1	4	3	1	2		3			

Table 5 shows a clear tendency for evaluators to focus on and utilize the first three of the ten heuristics. In fact, with one exception, all evaluators discovered at least one usability problem for each of the first three heuristics whereas the latter heuristics captured fewer problems. Examining the data more closely, we found that the evaluators could also experience challenges in distinguishing between the heuristics, in particular some confusion related to the difference between (1) visibility of system state and (6) recognition rather than recall.

Analysing the qualitative nature of some of the usability problems, we discovered a number of interesting findings that revealed the added value of testing in context. In the HW condition, temporal issues largely related to system performance and data input, e.g. "*slow to respond - took almost half the time necessary - would have been*

[1] In a small number of cases problems were reassigned to a more appropriate heuristic by the authors when it was clear that a better fit existed.

quicker with pen and paper". In contrast, the CW evaluators discussed temporal issues in terms of contextual pressures, e.g. *"There is a severe problem because the friend is waiting to leave the pub"*. No HE evaluators reported issues that related to time. We may speculate that situating the evaluation provides more traction over the less static and concrete dimensions of interaction.

Other evidence for the influence of context relates to movement and its impact on data entry. The movement of the lift in the third scenario triggered problem reports for the CW evaluators, as they found it difficult to enter data into the PDA in a moving lift, e.g. *"lift movement caused entering keyboard commands hard to do"*. Additionally, for the CW, physical restraints in the context would sometimes generate problems not discovered in the controlled environment of a laboratory, e.g. *"…not enough light from device in some background lighting condition"*.

With respect to resources required, the HE took an average of 60 minutes to complete, the HW 90 minutes and the CW 150 minutes. Additionally, HE was harder to manage than HW, the HE evaluators requiring more guidance from the facilitator. The use of scenarios in the HW guided the usability evaluators. The extra resources required for the CW related largely to the need to move location (between for example, a cafeteria, an elevator and a public house) whilst performing the evaluation.

5 Discussion

Despite a relatively small data set, it appears that adding scenarios to HE increases problem coverage, particularly at the more severe end of the spectrum. This may be due to scenarios sensitizing the evaluator to goal related activity. In HE evaluators were allowed to interact with the mobile device in a more open-ended manner. It was observed that without an explicit objective (that of accomplishing the task of the scenario) evaluators in HE seemed to examine the mobile device in an abstract, technologically oriented way. HW evaluators examined in detail those functions of the mobile device that were cued by the scenarios. Unsurprisingly but encouragingly, scenarios helped to steer the evaluation in a way that explicitly exposed the evaluators to particular aspects of the interface; ways which can of course be made to reflect the interests of the design team at large.

The scenario component in the HW seemed to aid evaluators in drawing on their own experience of the activities, and related contexts; issues largely ignored by the heuristics as written. Evaluators in the HW were more clearly playing the dual role of user and expert in evaluating the mobile device. The scenarios that were provided in the HW influenced the evaluators to step through the evaluation in a way that mimicked intended end-user behaviour.

In a closer examination of the common usability problems identified by both HE and HW, we note that the same usability problems were described in subtly different language. In HE, the evaluators described usability problems in terms of the mobile device itself, with little reference to context. That is, they focused more on the technology, e.g. *"icons not intuitive"*, *"category's icon not clear"*. This finding is consistent with Muller's [9] criticism that HE is product-oriented. In contrast, the HW de-

scribed problems at a detailed level, more closely related to the tasks that were specified in the scenarios. The poor performance of HE in relation to major flaws and catastrophes (and indeed in relation to the reclassified benchmark flaws) is consistent with the findings in other literature which report that HE flaws tend to be dominated by minor issues [11].

Initially it appeared that the quantity and nature of the usability problems identified in the CW were similar to those identified in the HW, and being in-situ added little. However, looking at the descriptions of the usability problems closely, it is clear that the CW discovered more severe problems that related to keyboard input, ambient lighting and processing speed of the PDA. In the CW, all three usability evaluators identified difficulties with pointer input. Two of these evaluators ranked this usability problem as a usability catastrophe while one evaluator ranked it as a major problem. Only one evaluator in the HW identified the accuracy of keyboard as a usability problem, with a ranking of minor. Hence, the same usability problem could be given a higher ranking in the contextual walkthrough. This might be because HW evaluators were in a controlled environment, whilst evaluators performing the CW in the field were reactive to the environment around them. For example, all three CW evaluators experienced elevator movement and commented that this interfered with entering keyboard commands. Similarly, for the HW ambient lighting in the laboratory was quite suitable, whilst for the CW the bar in which the equivalent scenario was explored was dark and lighting changeable, revealing difficulties concerning icon visibility. Further, the HW and CW differed in their view of the temporal aspects of use, with the CW focusing on the drivers in the context, giving weight to the input inefficiencies identified by the HW.

Variability in the number of problems discovered was greater in the HW than the CW. Placing evaluators in the field exerted a levelling influence, perhaps providing a better sense of the realistic environment. This reduced the differences in how they interpreted and imagined the context of use. On the other hand, in the HW, context was provided via written scenarios. HW evaluators interpreted the context as described in the scenarios differently, thus impacting the number of usability problems identified in the HW. A participant in the CW commented that *"scenarios are limited to the level of the context...how much detail and to the background of person interpreting them"*. A CW participant found it easy to identify flaws in the field because the physical environment cued him on his past experiences of using similar mobile devices, in a way that scenarios appeared not to. He found that the field made it easier to visualize how he might use the device.

All evaluators in the CW criticized the heuristics for their insensitivity to context, *"heuristics are completely environment immune"*. We have illustrated two ways in which such immunity can be breached, for the better.

It would be an interesting topic for future research to extend and adapt the heuristics for mobile use, so that they capitalise on the opportunities presented by HW and CW to examine mobile use in context; heuristics that address ergonomic issues, physical and social context, and mobility in more general terms, would be useful contributions. Additionally, supplementing scenarios with richer descriptions of con-

text, especially its less tangible social and dynamical features, would be of great value.

6 Conclusion

Adding scenarios to HE was an attempt to provide usability evaluators with a sense of the context of use- to bridge the realism gap. We found that the scenarios not only helped evaluators discover more critical usability problems, but the focus on these problems also shifted from being product-oriented to use-oriented. Furthermore, conducting the evaluation in-situ revealed usability problems, such as input-output issues and problems with ambient lighting, not encountered in-vitro. Interestingly the perspective taken by the evaluators on some problems, e.g. temporal issues, shifted as a function of being in-situ. By putting evaluators in the field, it may also reduce the variation between them in interpreting and imagining the context as written in the scenarios, however the benefits of in-situ application of Heuristic Evaluation, if they can be confirmed in further studies, do not appear to outweigh the costs in terms of time, training or gaining access to appropriate contexts of use.

Acknowledgments. Thanks to all 11 usability experts who gave their time generously, to our colleague Jesper Kjeldskov for useful comments on an earlier draft, and to two useful reviews that led to improvements in the paper.

References

1. Abowd, G. and Mynatt, E. (2000) Charting past, present and future research in ubiquitous computing. ACM ToCHI, Vol 7, Issue 1, Pp 29-58.
2. Kjeldskov J. and Stage J. (2004) New Techniques for Usability Evaluation of Mobile Systems. International Journal of Human-Computer Studies (IJHCS) Elsevier (forthcoming 2004)
3. Dunlop, M. and Brewster, S. (2000), The Challenge of Mobile Devices for Human Computer Interaction. Personal and Ubiquitous Computing, Volume 6, Number 4. Pp 235 - 236
4. Greenberg, S., Fitzpatrick, G., Gutwin, C. and Kaplan S. (1999) Adapting the locales framework for heuristic evaluation of groupware. Proc OZCHI2000.
5. Johnson, P. (1998) Usability and mobility; interaction on the move. Proc First Workshop on Human-Computer Interaction with Mobile Devices.
6. Kjeldskov, J and Graham, C. (2003) A review of MobileHCI research methods. Proc MobileHCI 2003.
7. Kjeldskov J. and Skov M. B. (2003) Creating a Realistic Laboratory Setting: A Comparative Study of Three Think-Aloud Usability Evaluations of a Mobile System. In Proceedings of the 9th IFIP TC13 International Conference on Human Computer Interaction, Interact 2003. Zürich, Switzerland. IOS Press

8. Law, L. and Hvannberg E. (2002) Complementarity and convergence of heuristic evaluation and usability test: a case study of UNIVERSAL brokerage platform. Proc NordicCHI 2002.
9. Muller, M., McClard, A., Bell, B., Dooley, S., Meiskey, L., Meskill, J., Sparks, R., and Tellam, D. (1995) Validating an extension to participatory heuristic evaluation: quality of work and quality of work life. Proc CHI'95.
10. Nielsen, C (1998) Testing in the field. Proc APCHI 98, pp285-290
11. Nielsen, J. (1994) Heuristic Evaluation, in Nielsen and Mack Usability Inspection Methods. Wiley.
12. Po, Shirlina, (2003) Mobile Usability Testing and Evaluation. BIS(Hons) Thesis. Department of Information System. The University of Melbourne.
13. Vetere, F., Howard, S., Pedell, S., and Balbo, S., (2003) Walking through mobile use: novel heuristics and their application. Proc OZCHI2003.
14. Nielsen, J. Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. Available at http://www.useit.com/papers/guerrilla_hci.html (reviewed 25/02/2004 1:13:30 PM)
15. Millen, D. R., & Dray, S. (1999). Information sharing in an online community of journalists. In Proceedings of Ethnographic studies in Real and Virtual Environments: Inhabited Information Spaces and Connected Communities. Edinburgh, Scotland

Is It Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field

Jesper Kjeldskov, Mikael B. Skov, Benedikte S. Als, and Rune T. Høegh

Aalborg University, Department of Computer Science
Fredrik Bajers Vej 7E, DK-9220 Aalborg East, Denmark
{jesper, dubois, als, runethh}@cs.auc.dk

Abstract. Evaluating the usability of mobile systems raises new concerns and questions, challenging methods for both lab and field evaluations. A recent literature study showed that most mobile HCI research projects apply lab-based evaluations. Nevertheless, several researchers argue in favour of field evaluations as mobile systems are highly context-dependent. However, field-based usability studies are difficult to conduct, time consuming and the added value is unknown. Contributing to this discussion, this paper compares the results produced by a laboratory- and a field-based evaluation of the same context-aware mobile system on their ability to identify usability problems. Six test subjects used the mobile system in a laboratory while another six used the system in the field. The results show that the added value of conducting usability evaluations in the field is very little and that recreating central aspects of the use context in a laboratory setting enables the identification of the same usability problem list.

1 Introduction

In the proceedings of the first workshop on Human-Computer Interaction for Mobile Devices in 1998, researchers and practitioners were encouraged to investigate further into the criteria, methods, and data collection techniques for usability evaluation of mobile systems [8]. Of specific concerns to the development of such methods and techniques, it was stated that traditional usability laboratory setups would not adequately be able to simulate the context surrounding the use of mobile systems and that evaluation techniques and data collection methods such as think-aloud, video recording or observations would be extremely difficult in natural settings. These concerns have since been confirmed through a number of studies e.g. [5, 6, 7, 9, 16, 18].

In 2003, a literature study on mobile HCI research methods revealed that 41% of mobile HCI involved evaluation [10]. However, even though evaluations of mobile systems are prevalent, surprisingly little research has been published concerning the methodological challenges described above. Exceptions include studies comparing methods applied for evaluating mobile systems in e.g. [5, 7, 9, 11, 17]. Consequently, no agreed upon set of appropriate usability evaluation methods and data collection techniques yet exists within the field of mobile HCI.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 61–73, 2004.

While the literature study [10] also revealed that 71% of mobile device evaluation was done through laboratory experiments and only 19% through field studies, it seems implicitly assumed that usability evaluations of mobile devices *should* be done in the field [1, 5, 8]. But field-based usability studies are not easy to conduct. They are time consuming and the added value is questionable [6, 9, 14, 16, 18]. Motivated by this, it has been suggested that instead of going into the field when evaluating the usability of mobile devices, requiring mobility or adding contextual features such as scenarios and context simulations to laboratory settings can contribute to the outcome of the evaluation while maintaining the benefits of a controlled setting [4, 9, 11, 12, 17, 20].

More emerging mobile systems are being characterized as context-aware as *they incorporate the ability of an application to discover and react to changes in the environment* [21]. Abowd and Mynatt state that the strong link to the physical context of a context-aware mobile systems challenge even further the conductions of usability evaluations as the *scaling dimensions that characterize context-aware systems makes it impossible to use traditional, contained usability laboratories* [1]. They continue by stating the effective usability evaluations require realistic deployment into the environment of expected use [ibid.]. However, we still have little knowledge about the relative strengths and weaknesses of laboratory-based versus field-based usability evaluations of context-aware mobile systems.

This paper has two purposes. Firstly, we want to compare the outcome of evaluating the usability of a mobile system in a laboratory setting and in the field in relation to identified usability problems and time spent on conducting the evaluations. Secondly, we want to describe two techniques used for 1) improving the realism of laboratory settings by including mobility and context, and 2) supporting high-quality video data collection when evaluating usability of mobile devices in the field.

2 Experimental Method

To address the above issues, we conducted a study involving two usability evaluations of a context-aware mobile electronic patient record (EPR) system prototype. The two evaluations involved a total of 12 professional nurses as test subjects conducting standard morning work routine activities. The first evaluation took place in a state-of-the-art usability laboratory where the subjects performed a series of assigned tasks while thinking-aloud. The second evaluation took place at the Hospital of Frederikshavn involving real work activities.

Studying usability of mobile systems for hospital settings takes the challenges of laboratory and field evaluations to an extreme. The users are typically highly mobile and the work pace is often intense and stressful. Furthermore, work activities are safety-critical (with errors potentially endangering the wellbeing or life of patients) and involve several ethical considerations such as privacy.

Recreating a healthcare context in a usability laboratory can be extremely difficult even impossible as such healthcare contexts integrate very complex work procedures, work situations, and tools. Recent studies on the usability of mobile information sys-

tems in healthcare have employed an indirect approach to data-collection about usability through interviews and questionnaires about user-friendliness and user-satisfaction with a prototype system [2, 23]. While overcoming some of the challenges described above, this approach does not provide first-hand insight into user-interaction.

2.1 The Context-Aware Mobile System Evaluated: MobileWard

Based on evaluations of stationary electronic patient record (EPR) systems and field studies of mobile work activities in hospitals, we implemented a context-aware mobile EPR prototype called MobileWard [22]. MobileWard runs on a Microsoft PocketPC based Compaq iPAQ 3630 connected to an IEEE 802.11b wireless TCP/IP network. The system was programmed in Microsoft embedded Visual Basic 3.0.

MobileWard is designed to support planning and conducting work tasks during morning procedure at a hospital department. The system is context-aware in the sense that the system presents information and functionality adapted to the location of the nurse, the time of the day, and the conditions of the patients. Based on the classification by Barkhuus and Dey [3], MobileWard is an *active* context-aware system as it automatically presents information and adapts to the context.

Before visiting assigned patients for morning procedure, nurses often want to get an overview of the specific information about each patient. As this typically takes place at the nurse's office or in the corridor, the system by default displays the overall patient list (figure 1a). Patients assigned for morning procedure are shown with a white background and the names of patients assigned to the nurse using the system are boldfaced (e.g. "Julie Madsen"). For each patient, the patient list provides information about previous tasks, upcoming tasks and upcoming operations. The indicators TP (temperature), BT (blood pressure) and P (pulse) show the measurements that the nurse has to perform. "O" indicates an upcoming operation (within 24 hours), which usually requires that the patient should fast and be prepared for operation. At the top of the screen, the nurse can see their current physical location (e.g. "in the corridor").

Fig. 1. MobileWard: Three different screen layouts of the context-aware mobile EPR system

The window in figure 1b displays information related to one patient including name and personal identification number of the patient, previous sets of measured tempera-

tures, blood pressures, and pulses as well as notes regarding the treatment of the patient. To enter new data into the system, the nurse must scan the barcode identification tag on the patient's wristband using the "scan" function in the bottom of the screen. When the nurse enters a ward, the system automatically displays information and functionality relevant to this location (figure 1c). Information about the patients on the current ward is presented, resembling the information available on the patient list displayed in the corridor, with the addition of a graphical representation of the physical location of the patient's respective beds. Data on each patient is available by clicking on the names.

In the evaluated prototype of MOBILEWARD, some of the contextual sensing functionality was simulated by means of a "context control centre" application. The control centre runs on a separate iPAQ connected to the wireless network. Through this application, an operator can trigger "context events" in MOBILEWARD, e.g. instructing the system that the user has entered a specific room.

2.2 Laboratory Evaluation

The idea of the laboratory evaluation was to evaluate MOBILEWARD in a controlled environment where we could closely monitor the use of the system. In addition to this, we also wanted to extend the standard experimental setup to include mobility and context. In order to achieve this, we modified the standard laboratory setup in a number of ways. The laboratory evaluation is described in detail below.

Setting. The usability laboratory was set up to resemble a part of the physical space of a hospital department (figure 3 and 4). This included the use of two separate evaluation rooms connected by a hallway. Each of the evaluation rooms were furnished with beds and tables similar to real hospital wards. From a central control room, the evaluation rooms and the hallway could be observed through one-way mirrors and via remotely controlled motorized cameras mounted in the ceiling.

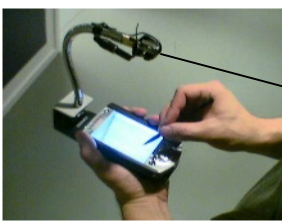


Fig. 2. Wireless camera mounted on PDA

Fig. 3. Video images from furnished subject rooms

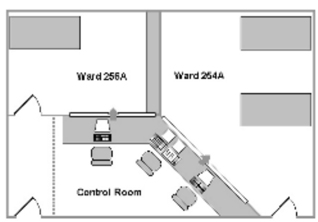


Fig. 4. Physical layout of the usability laboratory

Data collection. High quality audio and video data from the laboratory evaluation was recorded digitally. A tiny wireless camera was clipped on to the mobile device (figure 2), providing us with a close-up view of the screen and user-interaction. This was then merged with the video signals from the ceiling-mounted cameras (figure 2).

Test subjects. Six test subjects (four females and two males) aged between 28 and 55 years participated in the study. All test subjects were trained nurses employed at a

large regional hospital and had between 2 and 36 years of professional experience. They were all mobile phone users but only one had experience with the use of handheld computers. All test subjects were familiar with stationary electronic patient record systems and described themselves as experienced or semi-experienced IT users.

Tasks. All test subjects were given a series of tasks to solve while using the system. The tasks were derived from a field study at a hospital ward and covered the duties involved in conducting standard morning work routines. This primarily involved 1) checking up on a number of assigned patients based on information in the system from the previous watch, 2) collecting and reporting scheduled measurements such as temperature, blood pressure, and pulse, and 3) reporting anything important for the ongoing treatment of the patients should be taken into consideration on the next shift.

Procedure. Before the evaluation sessions, the test subjects were given a brief instruction to the system. This included the room-sensing functionality and the procedure for scanning patients' bar-code tags. The test subjects were also instructed on how to operate the available instruments for measuring temperature, blood pressure and pulse. The evaluation sessions were structured by the task assignments. The tasks required the test subjects to interact with all three patients in the two hospital wards, and move between the two rooms through the connecting hallway a number of times. The nurses were encouraged to think aloud throughout the evaluation explaining their comprehension of and interaction with the system. The evaluations lasted between 20 and 40 minutes and were followed by the test subjects filling out a questionnaire.

Roles. Each evaluation session involved six people. One nurse used the system for carrying out the assigned tasks. Three students acted as hospitalized patients. One researcher acted as test monitor and asked questions for clarification. A second researcher operated the context-control centre and the video equipment.

2.3 Field Evaluation

The second evaluation took place at the Hospital of Frederikshavn. The aim of this evaluation was to study the usability of MOBILEWARD for supporting real work activities at a hospital involving real nurses and real hospitalized patients. In order to achieve this, we adopted an observational approach combined with questions for clarification while the nurses were not directly engaged in conducting their work. The field evaluation is described in detail below.

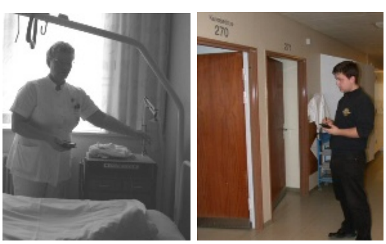


Fig. 5. Field evaluation at the hospital

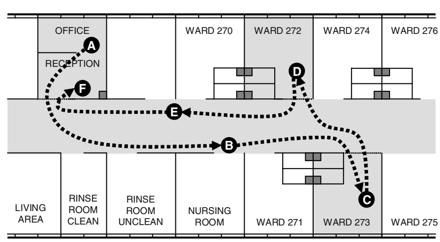


Fig. 6. Physical layout of the hospital wards

Setting. The field evaluation was carried out at the Medical Department at the Hospital of Frederikshavn (figure 5 and 6). This included the physical area of seven hospital wards, an office with reception, a rinse room and a break-out area connected by a central hallway and involved nurses at work and patients committed to the hospital.

Data collection. Motivated by the challenges of capturing high-quality video data during usability evaluations in the field, we designed a portable configuration of audio and video equipment to be carried by the test subject and an observer, allowing a physical distance of up to 10 meters between the two. The configuration consists of a tiny wireless camera (also used in the laboratory evaluation described above) clipped-on to the mobile device (figure 2) and a clip-on microphone worn by the test subject. Audio and video is transmitted wireless to recording equipment carried by the observer (figure 7). In the test monitor's bag, the video signal from the clip-on camera can be merged with the video signal from a handheld camcorder (Picture-in-Picture) and recorded digitally. This allows us to record a high-quality close-up view of the screen and user-interaction as well as an overall view of user and context. During the evaluation, the observer can view the user‘s interaction with the mobile device on an small LCD screen and monitor the sound through earphones.

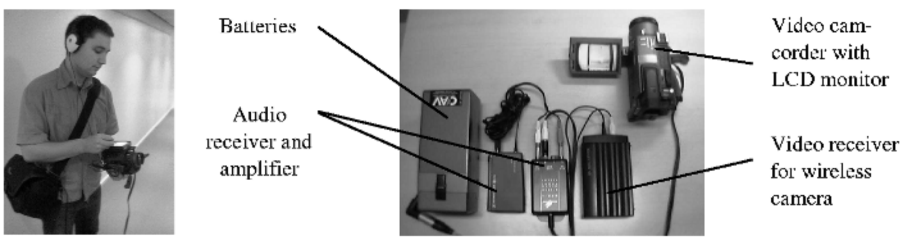


Fig. 7. Observer (left) carrying and operating portable audio/video equipment (right) for capturing high-quality data in the field.

For ethical reasons, we were not permitted to film the hospitalized patients allowing only the video signal from the clip-on camera to be recorded.

Test subjects. Six test subjects (all females) aged between 25 and 55 years participated in the field evaluation. All test subjects were trained nurses employed at the Hospital of Frederikshavn and had between 1 and 9 years of professional experience. They were all mobile phone users but novices with the use of handheld computers. All test subjects were frequent users of a stationary electronic patient record system and described themselves as experienced or semi-experienced users of IT.

Tasks. The field evaluation did not involve any researcher control in form of task assignments but was structured by the work activities of the nurses in relation to conducting standard morning work routines. As in the task assignments of the laboratory evaluation, the work activities of the nurses involved 1) checking up on a number of assigned patients, 2) collecting and reporting scheduled measurements, and 3) reporting anything important for the ongoing treatment of the patients.

Procedure. As in the laboratory evaluation, the test subjects were given a brief instruction to the MOBILEWARD system, including the room-sensing functionality and

the procedure for scanning a patient's bar-code tag. The evaluation sessions were structured by the work activities of the nurses which involved interaction with three patients in different wards and moving between different rooms through the connecting hallway a number of times. The nurses were encouraged to think aloud when possible. The evaluations lasted 15 minutes on average and were followed by the test subjects filling out a brief questionnaire. In order to be able to include a suitable number of nurses, the field evaluation took place over two days.

Roles. Each evaluation session involved six people. One nurse used the system for carrying out her work activities. One researcher acted as test monitor and asked questions for clarification while in the hallway. A second researcher operated the context-control centre application and the portable audio/video equipment. In addition, each evaluation session involved three hospitalized patients in their beds. Due to the real-life nature of the study, each evaluation session involved different patients.

2.4 Analysis

The data analysis aimed at creating two lists of usability problems identified on the background of the two experimental settings. The usability problems were classified as cosmetic, serious or critical based on the guidelines provided by Molich [13]. The two usability evaluations amounted to approximately 6 hours of video recordings depicting the 12 test subject's use of the system. All sessions were analyzed in random order by two teams of two trained usability researchers holding Ph.D. or Master Degrees in Human-Computer Interaction. Each team analyzed the videos in a collaborative effort allowing immediate discussions of identified problems and their severity (as adapted in [11]). As a guideline for the collaborative analysis, each identified usability problem would be discussed until consensus had been reached. The two teams produced two lists of usability problems. Subsequently, these two lists were merged into one complete list. Again, this was done in a collaborative effort, discussing each problem and its severity until consensus had been reached.

Resources spent on planning and conducting the laboratory and field evaluation respectively was calculated on the basis of a time log kept by the involved researchers.

3 Results

We identified a total of 37 different usability problems from the 12 laboratory and field sessions where eight problems were assessed to be critical, 19 problems were assessed to be serious, and ten problems were assessed to be cosmetic (see table 1).

Our study showed that the laboratory setting revealed more usability problems than the field setting. The six test subjects in the lab experienced 36 of the 37 usability problems whereas the six test subjects in the field setting experienced 23 of the 37 usability problems; this difference is extreme significant according to a Fischer's exact test ($p<.001$). 14 usability problems (1 critical, 9 serious, 4 cosmetic) were unique to the lab setting, whereas one serious usability problem was unique to the field.

Table 1. Distribution of total numbers of identified usability problems

	Laboratory (N=6)	Field (N=6)
Critical (N=8)	8	7
Serious (N=19)	18	10
Cosmetic (N=10)	10	6
Total (N=37)	36	23

Regarding the critical problems, the lab setting identified all eight critical problems and the field setting identified seven critical problems; this difference is not significant. Considering the serious problems, we find that the lab identified eight additional problems compared the field and this difference is strong significant ($p<.01$) whereas the difference in cosmetic problems is significant ($p<.05$).

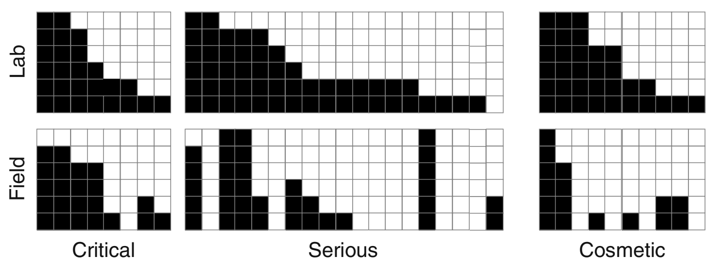


Fig. 8. The distribution of identified usability problems in the laboratory and in the field. Each column represents one usability problem associated the number of test subjects experiencing the problem (indicated by black boxes) for both settings.

Figure 8 summarizes the distribution of the identified 37 usability problems where each column represents one usability problem associated the number of test subjects experiencing the problem (indicated by black boxes) for both settings. Seven usability problems (two critical, two serious, three cosmetic) were experienced by all six subjects in the lab setting whereas three usability problems (two serious, one cosmetic) were experienced by all six subjects in the field setting; one usability problem (cosmetic) was experienced by all 12 subjects.

Looking across the distribution of the usability problems (in figure 8), we find that while the critical problems have a roughly similar distribution, the serious and cosmetic problems have rather dissimilar distributions where some problems were identified by all or nearly all subjects in one setting, but only identified by a few or none in the other setting. E.g. all subjects were informed to use either their fingers or the attached pen for device interaction, but only the lab subjects chose to use the pen and most of them experienced difficulties in placing the pen between tasks.

Analyzing the average numbers of usability problems identified per usability session, we find that the lab subjects on average experienced 18.8 usability problems (SD=2.0) and the field subjects on average experienced 11.8 usability problems

(SD=3.3) and this difference is strong significant according to a Mann-Whitney U-test (t=2.651, p<.01). This is mainly explainable through higher average numbers of identified serious and cosmetic usability problems where the difference of identified serious problems is strong significant (t=2.79, p<.01) and so is the difference of cosmetic problems (p=2.84, p<.01). On the other hand, we found no significant difference between the numbers of identified critical usability problems. This perspective on our data supports the findings illustrated in table 1 on total number of problems identified by six subjects in each configuration.

Table 2. The average number of identified problems per test session (standard deviations in parentheses).

	Laboratory (N=6)	Field (N=6)
Critical	5.3 (1.2)	4.5 (2.2)
Serious	7.5 (1.0)	4.5 (0.8)
Cosmetic	6.0 (0.9)	2.8 (1.0)
Total	18.8 (2.0)	11.8 (3.3)

Summarizing the time logs, a total of 34 man-hours were spent on the laboratory evaluation. Roughly 50% of this time was spent on planning the evaluation and setting up the lab while the remaining 50% was spent on conducting the evaluation sessions. In comparison, the field evaluation required a total of 65 man-hours. While the actual evaluation sessions in the field took less time than in the lab, the difference between the two was mainly accounted for by larger overhead for planning and transport and by more time spent on setting up the portable AV equipment and configuring MOBILEWARD with real data.

4 Discussion and Conclusions

The aim of our study was to identify opportunities and limitations of usability evaluation in laboratory and field conditions. Based on the results above; the numbers of identified problems, the nature of the identified problems and the lessons learned from conducting the two evaluations, we present the following four key findings:

(i) *Little added value of taking the evaluation into a field condition*. Quite surprisingly, our study shows that when compared to setting up a realistic laboratory study evaluators achieve very little added value when taking a usability evaluation of a context-aware mobile device into the field. In fact, in our study the laboratory setting was able to identify the exact same problems as in the field except for only one. This particular problem was related to an uncertainty expressed amongst some of the nurses at the hospital about the validity of data entered into the system, and whether it had been correctly saved in the database. The identification of this issue in the field

relates to the evaluation taking place during real work in a safety-critical use-context where errors cannot be tolerated. The fact that it was *only* identified in the field somewhat indicates a lack of realism in the laboratory condition.

The lack of added value of field evaluations contradicts the assumptions of more mobile HCI research studies, cf. [1, 5, 8, 14]. Here the general assumption is that evaluation of mobile, context-dependent and nomadic software should be conducted in their natural habitat in order to generate appropriate findings. In practice, however, this assumption is not taken into account by most research studies on mobile HCI as these typically apply laboratory-based evaluations [10]. Our results indicate that this may not be such a huge problem after all, and that expensive time in the field should perhaps not be spent on usability evaluation if it is possible to create a realistic laboratory setup including elements of context [11, 12, 20] and requiring mobility [9, 17]. As in the case of the evaluated system, field studies may instead be more suitable for obtaining insight needed to design the system right in the first place. Our results furthermore show that recreating the use context in a usability laboratory, as e.g. outlined by Nielsen [14], can produce successful mobile system usability results.

(ii) *Lack of control undermined the extendibility of the field condition.* Our study showed that the lack of control in field-based evaluations makes it challenging for evaluators to conduct field evaluations in practice and to make sure that every aspects of the system is covered. In our case, none of the field subjects used the note taking facility of the MobileWard system, leaving no chance for identifying usability problems in this particular system component. As we chose to have the actual work activities of the nurses directing the evaluation, we had no opportunities to force the use of the note taking functionality. This partially influenced the significant higher number of identified serious usability problems in the lab condition.

Issues often discussed in the usability literature are usability problem relevance and validity [13, 15, 19]. Artificial based evaluations e.g. think-aloud protocols in laboratory evaluations or heuristic evaluations may generate false positive problems that are not really problems in everyday use [13]. As a consequence, the higher number of identified problems in the lab condition could be a result of irrelevant usability problems; problems, which nurses would never experience when using the system in real life. However, our data does not exhibit whether this was the case or not. Finally, our field study was much more time consuming as it involved more preparation and travel cost; this is in line with findings of other research studies [6, 9, 14, 16, 18].

(iii) *Both the lab and the field revealed context-aware related problems.* For this particular study, we explicitly stressed the importance of context as the evaluated system was context-aware. Consequently, we would expect that in-situ evaluation could provide a different and perhaps more rich outcome. However, this was only vaguely the case. Both conditions identified all seven context-aware related problems, e.g. the problem of automatically updating information and functionality on the screen according to physical location was not always wanted by the subjects. Typically, they would either get confused or annoyed.

Surprisingly, however, all six field test subjects (but only one lab subject) got confused or did not understand why the system would automatically update information and functionality according to the physical location. So even though their use situa-

tion was in-situ and closely related to the context, they would still get confused of the system being actively context-aware (as defined by Barkhuus and Dey [3]). Analyzing this result, we find that their reluctance towards the automatic-update element in the mobile device may stem from the consequently decreased lack of control. Operating and working in a safety-critical environment like healthcare, the decreased level of control may not appear to support systematic work practices, but merely to compromise the work activities. The feeling of lack of control is well-known to active context-aware mobile system [3] and should probably be investigated further. Summarized, for professional work activities, our results seem to contradict statements from other literature on where to conduct evaluations of mobile systems, e.g. [1].

(iv) *The clip-on camera facilitated high-quality data collection of mobile use*. Aspects of mobility and of use in field settings typically challenge evaluators' opportunities for capturing the interaction between the user and the system. However, our configuration with a wireless device clip-on camera allowed the capturing of high resolution images of the interaction, which was invaluable during the later data analysis. The mobile configuration allowed the subjects to move freely in the environment, i.e. the lab and the field, while at the same time still providing us with the opportunity to record a close-up view of the interaction. The portable configuration of audio/video equipment made it possible to capture this data in the field.

Other studies have also stressed the importance of capturing the user-interaction and screen images of the system being evaluated [19], Generally, this have been found to be very difficult during mobile use [6, 8, 9]. Another way of dealing with this problem is to replicate screen images from the mobile devices on a laptop or stationary computer via a network connection and grab the images from here. However, this does not allow the capturing of situations where e.g. input is not registered by the system and does allow observation of user-interaction with the physical device. In laboratory settings, stationary cameras can be used to capture the screen of mobile devices too, but this approach is very sensitive to physical movements and typically requires the device to be held within a delimited area. In the field, video data is typically recorded by an observer with a handheld camera, continuously shifting focus between the mobile device, the user and the surrounding environment. However, this approach does normally not provide a very good view of the mobile device screen and user-interaction. Also, it requires the camera-operator to be in close proximity of the user and is highly sensitive to physical movement (which is, of course, prevalent during mobile use in the field).

Our study suffers from a number of limitations. First, the evaluated EPR system and the associated healthcare context probably influence the results of the study. Other domains may exhibit different characteristics where the link between the use of the system and the context may be weaker or stronger. Secondly, usability evaluations as applied in this paper provide only snapshots of intended future use. Other methods for understanding use and interaction like ethnographic studies can most likely provide different perspectives on context-aware mobile systems use. This could then in hand supplement or contradict the findings of our study.

References

1. Abowd, G. and Mynatt, E. (2000) Charting past, present and future research in ubiquitous computing. ACM Transactions on Computer-Human Interaction, Vol. 7(1), pp. 29-58
2. Bardram, J., Kjær, T. A. K., and Nielsen, C. (2003) Supporting Local Mobility in Healthcare by Application Roaming among Heterogeneous Devices. In Proceedings of Mobile HCI 2003, Udine, Italy, Springer-Verlag, LNCS
3. Barkhuus, L. and Dey, A. (2003) Is Context-Aware Computing Taking Control away from the User? Three Levels of Interactivity Examined. In Proceedings of the UbiComp2003 conference, LNCS 2864, pp. 149 - 156
4. Bohnenberger, T., Jameson, A., Krüger, A., and Butz, A. (2002) Location-Aware Shopping Assistance: Evaluation of a Decision-Theoretic Approach. In Proceedings of Mobile HCI 2002, Pisa, Italy, Springer-Verlag, LNCS
5. Brewster S. (2002) Overcoming the Lack of Screen Space on Mobile Computers. Personal and Ubiquitous Computing, 6, 188-205
6. Esbjörnsson M., Juhlin O. and Östergren M. (2003) Motorcyclists Using Hocman – Field Trials on Mobile Interaction. In Proceedings of the 5th International Mobile HCI 2003 conference, Udine, Italy, Springer-Verlag, LNCS
7. Graham R., Carter C. (1999) Comparison of Speech Input and Manual Control of In-Car Devices while on-the-move. In Proceedings of the Second Workshop on Human Computer Interaction with Mobile Devices, Mobile HCI 1999, Edinburgh, Scotland.
8. Johnson P. (1998) Usability and Mobility; Interactions on the move. In Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices, Glasgow, Scotland, GIST Technical Report G98-1
9. Kjeldskov J. and Stage J. (2004) New Techniques for Usability Evaluation of Mobile Systems. International Journal of Human-Computer Studies (IJHCS), 60, 599-620, Elsevier
10. Kjeldskov J. and Graham C. (2003) A Review of Mobile HCI Research Methods. In Proceedings of the 5th International Mobile HCI 2003 conference, Udine, Italy, Springer-Verlag, LNCS
11. Kjeldskov J. and Skov M. B. (2003) Creating a Realistic Laboratory Setting: A Comparative Study of Three Think-Aloud Usability Evaluations of a Mobile System. In Proceedings of the 9th IFIP TC13 International Conference on Human Computer Interaction, Interact 2003. Zürich, Switzerland. IOS Press
12. Lai J., Cheng K., Green P., Tsimhoni O. (2001) On the Road and on the Web? Comprehension of synthetic speech while driving. In Proceedings of CHI'2001, ACM, New York, pp. 206-212
13. Molich, R. (2000) *Usable Web Design* (In Danish). Ingeniøren | bøger
14. Nielsen C. (1998) Testing in the Field. In Proceedings of the third Asia Pacific Computer Human Interaction Conference, IEEE Computer Society
15. Nielsen J. (1993) Usability Engineering. Morgan Kaufmann
16. Pascoe J., Ryan N., Morse D. (2000) Using While Moving: HCI Issues in Fieldwork Environments. Transactions on Computer-Human Interaction, 7(3), 417-437.
17. Pirhonen A., Brewster S.A., Holguin C. (2002) Gestural and audio metaphors as a means of control for mobile devices. In Proceedings of CHI'2002, ACM, New York.
18. Rantanen J., Impio J., Karinsalo T., Reho A., Tasanen M., Vanhala J. (2002) Smart Clothing Prototype for the Artic Environment. Personal and Ubiquitous Computing, 6, 3-16.
19. Rubin, J. (1994) Handbook of Usability Testing. Wiley

20. Salvucci D. D. (2001) Predicting the Effects of In-Car Interfaces on Driver Behaviour using a Cognitive Architecture. In Proceedings of CHI'2001, New York, pp 120-127
21. Schilit, B. N. and Theimer, M. M. (1994) Disseminating Active Map information to Mobile Hosts. In IEEE Network, Vol. 8(5), pp. 22-32
22. Skov, M. B. and Høegh, R. T. (2004) Supporting Cooperation and Coordination in a Hospital Ward by a Context-Aware Mobile Electronic Patient Record. Submitted to the Journal of Computer-Supported Cooperative Work, Elsevier.
23. Urban, M. and Kunath, M (2002) Design, Deployment and Evaluation of a Clinical Information System which uses Mobile Computers and Workstations. In Proceedings of the 2nd conference on Mobile Computing in Medicine, Heidelberg

Home Is Where Your Phone Is: Usability Evaluation of Mobile Phone UI for a Smart Home

Tiiu Koskela[1], Kaisa Väänänen-Vainio-Mattila[1], and Lauri Lehti[2]

[1] Institute of Software Systems
{Tiiu.Koskela, Kaisa.Vaananen-Vainio-Mattila }@tut.fi
[2] Institute of Electronics
{Lauri.Lehti}@tut.fi
Tampere University of Technology,
P.O.Box 553, FIN-33101 Tampere, Finland

Abstract. Many home activities are mobile within the home. Now, in the beginning of the era of smart homes, the mobility of home activities can be taken even further. Networking technologies enable inhabitants of a smart home to access home functions from a distance. A mobile device, in specific the mobile phone, can act as a "remote control" UI for the smart home functions. This paper presents the results of research in which the usability and acceptability of a mobile phone as an UI to smart home functions were evaluated. Evaluation was done through focus groups, usability tests of a mobile UI prototype and a three-month usage experience of a young couple living in a smart apartment. The results indicate that the mobile phone is especially an attractive UI when *instant control* of both within-home and over-the-distance functions is needed. Furthermore, users liked the increased convenience and feeling of safety as the mobile phone enabled them to "feel home" from the distance.

1 Introduction

In the home environment the inhabitants' tasks form patterns or chains of actions. Individual actions scatter among different *action centers* [4]. Thus the nature of everyday tasks is often rather mobile; the task-performance is started at one action center and continued at another. For example the mail can be distributed to the family members via the hallway table; the important ones, like bills and invitations can be attached on the fridge door in the kitchen and the personal post is taken to the bedroom to the bedside table to wait for a private moment. In addition home is a multitasking environment, where switching between different actions and activity levels is common. Dinner, for example, is cooked while the cook is engaged in a telephone conversation. Thus interaction with smart home functionalities should enable mobility and interruptions. The interaction should form a sequence of steps that can be resumed and built upon [1].

The requirement for mobility rises from the needs for mobile control *inside* the home but also from the increasing needs for remote control from *outside* the home. The

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 74–85, 2004.

need for control home while away arises from the small changes in how people experience home. For example, virtual communities are shaping the concept of home [5]. Home can be seen to grow over the specific physical place and location. A smart home should not be seen as being restricted strongly to only one place – house or an apartment – rather it is a *way of life*. Advanced mobile computing enables people to have *continuous presence* or connection to selected people and services, some of which may be found in the domestic context [11].

Smart home research has focused on developing *ubiquitous computing* solutions, which aim to adjust to the inhabitants' needs according to the information collected from the inhabitants, the computational system and the context. Ubiquitous computing enables invisibility of the computational technology and provides the user with "natural" interaction techniques, such as speech [1]. Nevertheless, it can be questioned whether people are ready for the invisible ubiquitous computing. According to Rodden et al [10] the future smart homes must evolve from our existing homes. In order for smart homes to gain social and practical acceptability, the new technology should intermesh with the old. The research of this evolution phase should focus on determining the roles of the familiar technologies and devices in the home environment – whether these objects can provide means for smart homes to evolve.

This paper presents the results of research where the mobile phone as a UI to smart home was evaluated. The usability and acceptability of the user interface prototype was analysed through focus groups, laboratory tests, and an ethnographic research of three-month usage experience in a real smart home environment. The special value of our research lies in its empirical settings, as in earlier research of others the end-user experience of smart home solutions has been studied via usability tests and walkthroughs of Wizard of Oz-type prototypes [3] and brief field trials [8]. The test subjects of longer trial periods have in many cases been members of the design team [6].

2 Smart Home Usability and Living Experience Research

The usability research of the mobile phone user interface presented here was a part of the Smart Home Usability and Living Experience project. The research was carried out during May 2002-March 2003 at the Institute of Software Systems of Tampere University of Technology (TUT) with support from Nokia Mobile Phones, Pikosystems (accessibility solutions), and Tekes (National Technology Agency of Finland). Institute of Electronics of TUT developed the smart apartment and the UI prototypes for the research purposes. The project investigated the usability and acceptability of interaction solutions for smart home environments.

2.1 Smart Home Test Environment

In order to be able to do empirical research of smart environment systems, a test apartment – the *eHome* – was set up in a new apartment building neighbourhood near

the center of the city of Tampere in southern Finland. This smart home was an apartment with a living room, bedroom, kitchen, bathroom and sauna.
A selection of test functionalities such as remote control of room lights and venetian blinds was implemented in the smart apartment. All the lights, curtains and electric devices could be controlled individually, and the lights could also be controlled in user-defined groups. The status of the plants and electric devices could be monitored. In addition, different kinds of timings of the devices could be created.

Inhabitants were able to interact with the smart functionalities and devices using three different kinds of user interfaces: a mobile phone, a laptop computer (with WLAN) and a TV with media terminal, which was used via the remote control of the TV[1]. The TV's remote control and the laptop computer were connected via intranet and the mobile phone was connected via Internet to the core server. The core server took care of the communication between the devices and the UIs. The sockets and lights were controlled via a LINET control system. All the network technology in the apartment was hidden from the inhabitants.

2.2 Research Process

The usability research process was adapted from the human-centered design process set by the ISO standard 13407. The research was carried out via three phases: *definition, design* and *evaluation.*

In *the definition phase* the basic user requirements for the user interface design were collected from literature, through contextual inquiry [2], theme interviews and focus groups [9]. The nature of everyday life patterns and needs for home technology were examined with the contextual inquiries and interviews in people's own homes. Focus group sessions were held in order to chart the subjects' attitudes and presumptions of the future homes. The subjects of the contextual inquiries, interviews, and focus groups (total of 22) had no previous experiences of smart home solutions. The subjects consisted of adults, families with children, middle-aged couples, and elderly people.

The design phase consisted of the design and implementation of the user interface prototypes, whose usability was examined via heuristic analyses and usability tests. The tests were carried out in lab with 5 test users in order to find at least 80% of the possible usability problems. The real end-users in the ethnographic evaluation phase (the inhabitants of the smart apartment) were to be young adults; therefore test users selected for this design phase were mainly young adults.

In *the evaluation phase* the mobile phone user interface was installed in the smart apartment, where the inhabitants could use them in their everyday lives. The empiri-

---

[1] This paper focuses on the mobile phone UI. The other two UIs have been discussed in detail in [7]

cal study of the actual living experiences was carried out using an ethnographic approach with methods like contextual inquiries and participatory walkthroughs during three months. During the test period the users could use all three user interfaces (media terminal, PC, mobile phone). The inhabitants were selected from the applicants for the apartment. The profile requirement was that the tenants should not be technically oriented, but also not strictly resistant to new technology. The selected tenants were a couple of a 26 year-old woman (history researcher) and a 27 year-old man (biologist).

3 User Requirements for Smart Home UIs

In the definition phase of this research, initial user requirements were gathered via contextual inquiries, theme interviews and focus group sessions. The aim was to examine the domestic context and especially to discover what could be acceptable and desirable interaction solutions for the smart home.

3.1 Activity Patterns at Home

Tasks at home vary according to their necessity (e.g. doing dishes vs. watching TV), frequency (e.g. preparing dinner vs. watering plants) and the level of mental or physical activeness they require (e.g. vacuuming vs. listening to the radio). Performance of tasks that are either obligatory or frequent generates *activity patterns* [4].

Everyday activity patterns became evident in this study. The purposes for the patterns are manifold: The patterns bring *efficiency* to the task performance but their purpose is also to create a *feeling of safety* and *belongingness*.

The subjects desired more time free from planned and obligatory tasks. Ironically, it seems that in order to have this increased free time, the daily life must be well organized. Patterns are formed to meet this requirement for *efficiency*. The patterns can consist of multitasking, where the focus is switched between different tasks. The level of activeness and the physical location for the performance changes through the pattern. Thus, there is a need for mobile and flexible ways for interaction with smart home functions.

Consistent activity patterns provide a *feeling of safety,* which enables the user to gain the control or at least the feeling of control. Some of the study subjects' main fear for smart home solutions was that they might lose the control. This could happen both in everyday level (e.g. the VCR is too complex to use) and especially in situations of malfunction or maintenance problems, where the familiar "wrench and hammer" methods cannot be used anymore (e.g. when the software for the door lock should be updated). Subjects desired easy-to-use, consistent and familiar interaction methods.

The *feeling of belongingness* is achieved by the sense of the family members' presence. In addition it can be achieved by forming activity patterns around the media which provide connectivity to communities. For example, the newspaper is read while eating breakfast, or the radio or TV is listened while dressing up, or chat sites are visited between other tasks. These devices are used for various reasons and purposes and they have an important but flexible role in the home environment.

3.2 Attitudes Towards Interaction Techniques

In order to assess the appeal of different interaction techniques, smart home usage scenarios were presented to the focus groups. The participants were asked about their opinions on possible new interaction techniques: speech, gesture, graphical user interface (GUI), and automation, in which the home adapts to the users' actions [3] [7].

The subjects ranked *automation* as the most wanted interaction technique. However, it should not be "full" automation but instead they wanted to be able to define the causalities themselves: *"When this alarm stops, I want that lamp to turn on."* The participants did not trust the computers to understand the context of home and that in some cases the practicalities are not so important as social or emotional issues.

The general interaction techniques did not totally satisfy the subjects. They wanted something that enabled mobility of the task performance, but in a familiar way. All of the focus groups ended up with a concept of *centralised remote control*. A remote control would be a preferable solution, because it provides the needed mobility, and the general concept has become quite familiar by e.g. TV controls and mobile phones.

The usage and needed amount of remote controls were discussed widely. The key questions were if there should be a specific remote for every room, and should every inhabitant have his/her own device. Especially the need for *personal* remote controls raised questions. On one hand, the question was about *immediate access* if everyone has their own device they can carry around. On the other hand, the subjects pondered whether every inhabitant should have access to all the smart home functionality. The groups came to different conclusions: The family group and the middle-aged group felt that there was no need for individual controls. The young adults' group thought they would prefer a personal control device or at least personal identification to the commonly used control. This reflects the fact that young people are more accustomed to personal computing.

The participants agreed that the usage of the control should be efficient and easy. In addition they pointed out the importance of feedback. They considered that an ordinary TV remote does not provide enough information of the possible actions or the success of the performed actions. The participants stated the ordinary TV remote to consist of too many buttons and too little instructions. The participants were also attracted to the idea of immediate visual feedback.

3.3 Needs and Requirements for Mobile UI

Based on the results of the focus groups and contextual inquiries of potential smart home inhabitants, Table 1 presents a summary of user requirements for the mobile UI.

Table 1. Summary of the user requirements for the mobile UI for smart home functions

No	Users needs/wants	Requirements to the mobile UI
1	Users want efficient task performance according to their own activity patterns, which may involve multi-tasking and varies according to the level of activeness.	Flexible and mobile interaction with interruptions enabled. The interaction should form a sequence of steps that can be resumed and built upon.
2	Users want to feel safe and be in control of their home environment.	The interaction methods should be easy to use, consistent and familiar.
3	Users want the home to be a place where they can belong and enjoy the presence and company of family members.	The interaction methods should support switching between privacy (e.g. personal interaction) and togetherness (e.g. see how the others have changed the status of the home).
4	Users want a centralized remote control to smart home functions.	The interaction method should provide a centralized way to interact with the home devices, and should not require the user to be present at the same location with the controlled device.
5	Users want a remote control that is as simple as possible.	The interaction should require the usage of only 2-3 buttons.
6	Users want the interaction method to guide them with the task performance.	The interaction methods should give information about the available actions and clear feedback of the results of users' actions. Visual information would be preferable.
7	Young users want personal and mobile means for controlling.	The interaction methods should be provided via a personal interaction device or identification of different users.

4 Mobile UI Prototype for Empirical Study

Based on the requirements from the definition phase, three functional user interface entities were prototyped: The mobile phone UI on Nokia 6310i, a GUI on a laptop PC, and a menu-based UI on a media terminal connected to the TV and used with a normal remote control of the TV. This section presents the mobile phone UI prototype and the results from the usability tests. The other two UIs are not described here in detail but the results of the comparison are discussed briefly in the end of this paper.

4.1 Mobile Phone UI Prototype

The mobile phone UI can be used as a centralised means (requirement 4) to monitor and control smart home devices both within the home and from outside of home (req. 7). The mobile phone UI is menu-based and follows the mobile phone UI style users are accustomed to (req. 2). The UI allows the user to navigate menu hierarchies with the arrow buttons on their phones (req. 5). The hierarchies provide the user the task performance in a sequence of steps (req. 1) with instructions and visual feedback of the options for user's actions. Only the functionalities that are possible and logical to use are visible in the view (req. 6). The mobile phone is considered a personal device, thus it supports mainly the controlling done by the individual users (req. 3).

The set of functionalities that was implemented in the smart home was meant to give the smart home users an experience of a good selection of different types of home activities. This ensured that support of the UIs to the different types of activity patterns could be tested. The smart home functionalities were relatively simple, because the main purpose of the research was to focus on the UI usability and not to test the acceptability of individual smart home functions. Devices that can be controlled with the mobile phone application are divided in four groups: lights, curtains, plants and electric devices (Figure 1).

Fig. 1. The main menu and the screens for controlling the curtains

The results from the focus groups and contextual inquiries indicated that smart home solutions are most acceptable when they focus on security ("*Did I leave the coffee maker on?")* and energy saving issues ("*Are the lights on even when they are not needed?").* The mobile phone UI provided the users with a means to monitor the state of electrical appliances by controlling all electric sockets of the smart home. If the electric device is forgot on it can be turned off (Figure 2).

Fig. 2. Controlling the electric devices: Turning off the coffee maker

The prototype provides controls for the lights in the living room, in the kitchen, in the bedroom, in the hall and in the entire home. The entire home means that all the rooms can be controlled at once.

One of the expectations of the smart home was to free the inhabitants from house work and to provide means to take care of the home from a distance, especially during holiday season. The mobile phone prototype provided monitoring of the status of the plants at home. The status information is based on lighting, room temperature and mould moisture. With the collected information users can estimate the conditions of the plants. For example, in case of lighting, it could be adequate or the room might be too dark, in which case the curtains can be opened. Ideally, room temperature should also be controllable but this was not possible at the time of this research.

The research participants (focus group, contextual enquiries) emphasised the need for the home to be flexible and to suit many purposes. The suitability is affected by practical issues but also emotional and social issues: The users should have the possibility to change the mood of the room, e.g. via lights or sounds (“*Oh, how hard it is to get up—if only the sunlight would ease the pain!*”). Controllable curtains are in the living room and bedroom. Curtains are venetian blinds that can be controlled between 0–180 degrees. With the mobile phone application curtains can only be closed or opened fully.

4.2 Results from Usability Tests

Five test users performed a usability test of the mobile phone UI in a lab. In addition to the test tasks, users were interviewed after the tests to gather their views and further ideas for smart home UIs.

The tests showed that the user interface was usable without any major problems. The users considered the structure of the UI hierarchy so clear and easy to use that they hoped the functionality to be increased. For example, one user said that she would probably be able to set the timer of the VCR on if the service were provided by the user interface in the mobile phone.

Even though the hierarchy used in the user interface was considered clear, alternative ideas were presented. The tests indicated that a structure that would be based on the floor plan of the home could also be intuitive to the users. The users could then imagine themselves being at home even though they are controlling the home from a distance. For example, when one test user was trying to find out in the test situation whether the plants needed to be watered, she made assumptions that the plants are in the living room or kitchen. The home is seen as an entity where everything has its own place.

Regarding the breadth of access to the smart home devices, the test users thought that the user interfaces should have different functionalities for children and adults. The

young adults (3/5) were stricter than the middle-aged subjects (2/5) who had children of their own. According to the middle-aged test users the limitations should be determined according to the limitations in the manual usage of devices and functionalities. For example, if a child is allowed to use a stove at home, she/he should be allowed to control it, for example turn it off, from a distance too. The young adults themselves feared that the usage could become too playful.

The test users thought it was important to occasionally enlarge the user group and modify the usage rights. For example, during a holiday the house watchers should be able to use the mobile interface. In addition, if a grandparent ended up in hospital the relatives should be able to control the home and make it seem inhabited to chase away burglars. A possibility to create profiles for different users would be needed.

The reliability of the user interface provided by the mobile phone concerned the usability test subjects. Control from a distance raised doubts about whether the power or device really was really switched off when that was indicated by the user interface. Nevertheless a mobile phone has a more alluring image than some other ordinary domestic appliances – such as the PC – due to its possibilities to be personalized and carried with. Regarding the wanted smart home functionalities, test users mentioned checking the status of any electrical devices, such as the iron, watering the plants, setting the VCR to record, putting the sauna on, putting the lights on, checking what is in the refrigerator, even putting the food into an oven and preparing a bath.

5 Living Experience Results

In the evaluation phase the mobile phone UI was used for three months in the eHome by the young couple living in it. This section presents the results of the actual living and usage experiences, which were researched through contextual inquiries and participatory walkthroughs of the UIs and the implemented smart home functionality.

5.1 Need for Instant Control While at Home

Some of the home activity patterns are predictable and can be determined via causality by time or context information. These recurring tasks could be controlled via *pattern control* [7], which would enable user-controlled automation ("*When this alarm sets off, I want this lamp to turn on.*"). But as the saying "my home is my castle" suggests, the inhabitant should be able to both make and break the rules. Thus there is a need for flexible, *instant control* of home functionalities (*"Turn the lamp off now, I have changed my mind."*). In the home environment the user is often very impatient, and the functionalities and controlling should be provided "right here and now". The user interfaces and devices should be in a *constant stand-by* state, the users should be able to perform simple tasks via only a few actions and the controlling should be possible in the usage context of that moment even though the actions would be mobile.

The research showed that graphical UI of the PC suited well for the *pattern control,* and the mobile phone user interface supported well the *instant control.* The advantage of mobile phone UI was that it was truly functional on its own, for example no pointing was needed to the direction of the smart devices (as in the usage of media terminal and its remote) and there was no need to sit down in front of a display and keyboard (as with the PC UI). Therefore, the mobile phone UI provided full mobility and made controlling possible just where and when the user wanted. Secondly, the user interface provided simple and fluent task-flow with only a few steps. The users thought the user interface to be easy to use even though they were not familiar with this usage purpose of the phone. Thirdly, the mobile phone UI created the illusion of being more efficient than the other user interfaces. The inhabitants kept the mobile phone always on. Therefore even though the phone application for the smart home had to be selected it seemed to the users that they were immediately able to start performing the task. These aspects made the mobile phone user interface the most liked UI.

Before using the mobile phone UI, the inhabitants were prejudiced towards its usage. The couple did not think they would use the mobile phone at home: *"This is such a small apartment that there is no need for mobility."* The users did not have a regular landline phone so the mobile was often within their reach, and this enabled easy access to the home functions. The UI provided especially some *luxury* to the inhabitants. For example they found it rather enjoyable to just lie in bed and turn off the lights with the mobile phone, instead of getting up and using the switches on the wall.

In the end of the evaluation phase the couple was of the opinion that there should be an even easier way for controlling the lights for example. They suggested that some functionality could be defined based on contextual data. For example, the lights could be controlled according to the information on whether the inhabitants are at home or not, or whether they are sleeping or not. In the couple's opinion this kind of automation does not limit the need to have also some kind of a remote control for the functionalities. The automation would be used mainly when the user would be in an inactive state (e.g. absent or sleeping), and the mobile UI in the more active situations.

5.2 Need to Ensure Secure and Pleasant Home from a Distance

The mobile phone is one of the essential items in addition to keys and money that the inhabitants took with them when they left home. This makes the mobile phone a natural user interface device for controlling smart home functionalities from a distance. In addition the role of the mobile phone is flexible. The users are accustomed to the fact that the new series or models will provide new additional functionalities. The mobile phone is already seen as a terminal via which connections to people and services can be made, and the extension towards home environment control seems plausible.

The inhabitants' motivation to access home from a distance concerned the security and pleasantness of the home. The control from outside of home focused mainly on

the functionalities of lights and curtains. A typical usage situation was that the users wondered whether the lights were left on or not. After checking they adjusted the lights if needed. Thus controlling from a distance can be based on the need to prepare the home for the absence of inhabitants, because many times the home is left in a hurry. Another typical usage situation was that the users wanted to make the home-coming more pleasant, so they would put the lights in the hallway on while they were still in the car on their way home. The home is perceived to be a place where one can relax. Thus the users want the controlling from a distance to focus on the aspects that prepare the home for enjoyment, such as turning the sauna on (a frequent activity in Finland). According to the users there is also a need to access the information at home (e.g. stored in the home computer) from a distance. Home also has a role as a resource storage; home is a place where its inhabitants gain resources (rest, support and information). The home should become open for the authorised users' information retrieval but still remain restricted to others.

Even though control from a distance seemed appealing to users, it was not yet totally acceptable. The test inhabitants were not able to fully trust the technology. For example they considered turning on and off the electricity too risky to be used while they were not at home themselves.

6 Summary and Discussion

This paper reports the usability research of a mobile phone user interface to smart home functionality. The research was conducted in Tampere, Finland, in May 2002-March 2003. The project started with a definition phase in which user requirements were gathered from 22 subjects of different family statuses. Usability tests of the mobile phone UI were conducted prior to installing the UI to the eHome. An ethnographic research phase was conducted of the actual usage experiences by a young couple living in the smart apartment.

The research setting was based on the hypothesis that the smart home will evolve from current homes and thus the means for interaction should also be provided via familiar devices. The mobile phone was selected as one of the user interface devices, because due to its personal image, it was assumed to have a good potential to serve as a UI for centralised control. The other UI devices (PC and TV) were selected due to their flexible roles (as information resource, lamp, clock, etc.) in current homes.

The research confirmed the findings of Crabtree et al., that everyday actions in the home context form patterns [4]. In this study the patterns could be seen to create requirements for two diverse user interface types: UI for *pattern control* and UI for *instant control*. The pattern control consists of usage in which the user can predetermine the desired actions. This type of user-controlled automation can be well supported by the GUI on the PC. The mobile phone UI on the other hand suits well the instant control, which assumes immediate control of smart home functionalities

needed in that moment and context of use. In the home context the users are rather impatient and unwilling to wait; they should be able to perform the actions “right here and now”. The mobile phone UI enables this by providing the user a logical sequence of steps in familiar menu hierarchies. In addition, providing the user clear visual feedback and indications of the possible options for the user’s actions increases efficiency. This made the within-home interaction with the smart home functionalities attractive for the users in the ethnographic research phase.

The general tendency of people’s lives is towards mobility. The concept of home is psychologically extending outside the physical home. Therefore the control of home functions over distance can be used to provide efficiency, safety and feeling of control. The mobile phone can be seen as an extension of its users. The mobile way of life and the possibilities to be continuously connected with people and home functionalities enable the inhabitants of smart homes pleasantly “feel home” with the mobile phone also over a distance.

References

1. Abowd, B. D., Mynatt, E. (2000). Charting Past, Present, and Future Research in Ubiquitous Computing. *ACM Transactions on Computer-Human-Interaction*, vol. 7, No.1. ACM Press, NY, USA.
2. Beyer, H., Holtzblatt, K. (1998). *Contextual Design: Defining Customer-Centered Systems.* Academic Press, San Francisco, USA.
3. Brumitt, B, Cadiz, J (2001). “Let there be light”: Examining interfaces for homes of the future. *In Human Computer Interaction. INTERACT’01,* IOS Press, Amsterdam.
4. Crabtree, A., Hemmings, T., Rodden, T (2002). Pattern-based support for interactive design in domestic settings. *Proceedings of the conference on Designing interactive systems: processes, practices, and techniques.* London, England.
5. Elixmann, M. (2003). Life 21 – Mobile communities. *The new everyday. Views on Ambient Intelligence*. Ed. Aarts, E., Marzano, S. 010 Publishers, Rotterdam, The Netherlands.
6. Hindus, D, Mainwaring, S., Leduc, N., Hagström, A., Bayley, O. (2001). Casablanca: designing social communication devices for the home. *Proceedings of the SIGCHI conference on Human factors in computing systems.* Seattle, Washington.
7. Koskela, T., Väänänen-Vainio-Mattila, K. (2004). Evolution towards Smart Home Environments: Empirical Evaluation of Three User Interfaces. *Proceedings of 2AD: Second International Conference on Appliance Design,* May 11-13 2004, Bristol UK (Submitted to).
8. Mynatt, E. D., Rowan, J., Craighill, S., Jacobs, A. (2001). Digital family portraits: supporting peace of mind for extended family members. *Proceedings of the SIGCHI conference on Human factors in computing systems.* Seattle, Washington.
9. Preece, J., Rogers, Y., Sharp, H. (2002). *Interaction design: beyond human-computer interaction.* John Wiley & Sons, Inc, NY, USA.
10. Rodden, T., Benford, S. (2003). The evolution of buildings and implications for the design of ubiquitous domestic environments. *Proceedings on Human Factors in computing systems.* Florida, USA
11. Väänänen-Vainio-Mattila, K. (2002). The future of mobile communities: evolution towards continuous presence. *ACM SIGCHI Bulletin – a supplement to interaction,* May-June 2002, p. 9-9. Ed. Matt Jones.

User Validation of a Nomadic Exhibition Guide

Barbara Schmidt-Belz [1] and Fabian Hermann [2]

[1] Fraunhofer FIT,
53754 Sankt Augustin, Germany
Barbara.Schmidt-Belz@fit.fraunhofer.de
http://www.fit.fraunhofer.de/profil/icon_en.html

[2] Fraunhofer IAO,
70569 Stuttgart, Germany
Fabian.Hermann @iao.fraunhofer.de

Abstract. This paper reports on the user-oriented development of a nomadic exhibition guide for trade fair visitors. The system is situation aware, and personalized, supporting all phases from planning at home, over the mobile situation when visiting the exhibition, until the evaluation afterwards. The prototypical implementation of the system at a trade fair was basis for a final user evaluation. Users rated the importance of information retrieval, interactive, location aware maps and tour planning very high, while sophisticated features such as pro-active personalized tips, web-cam pictures, or locating colleagues were considered less important. Evaluation results concerning map visualization on small screens and egocentric navigation support are reported in more detail.

1 Introduction

Visiting a trade fair may feel like searching a jungle for some favourite fruits. The visitor to a trade fair is exposed to an almost aggressive mix of information presented in various media and spatially distributed over a vast exhibition site. Business people visiting a trade fair need to make best use of their time, focusing on their special interests and efficiently touring the exhibition. Many of them use the WWW to plan ahead, looking for information about companies, products and events. When touring the fair grounds, they use the exhibition catalogue, or also bring notes and printouts they have prepared in advance.

Today's technology is ready to realize a mobile exhibition guide that assists the user to find the information they are looking for and also guide them to the places they want to visit. A modern PDA has the necessary capacity, both storage and the computing, and also an adequate screen size and resolution, to serve as a mobile client for a trade fair guide. A variety of research prototypes of mobile guides have been realized over the last 10 years, for instance Cyberguide [1], Guide [2], Hippie [3], and CRUMPET [4]. While the science community now looks ahead to wearable, pervasive, and ambient computing, there is still a gap between the feasibility of mobile

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 86–97, 2004.

guides in principle, as demonstrated by research prototypes, and the fact that for most application domains (e.g. trade fairs and other exhibitions, tourism) the information services are actually on the level of WWW portals, catalogues on a PDA, and the occasional audio guide.

The SAiMotion project aimed to combine and further develop several computing and sensor technologies to realize an exhibition guide that could successfully go to market in the near future. The project explicitly chose a user-oriented approach in order to develop a system that actually meets user needs and has user-perceived added values compared to existing visitor information services at trade fairs.

For a set out, we assumed the guide should be

- *Nomadic*, i.e. it supports the user in a continuous way throughout all stages of a visit to the trade fair, beginning with the planning, during the mobile situation of use, and continuing after the visit when evaluating the results at home.
- *Context-aware*, including location-awareness in the mobile situation, adaptivity to personal interests, and adapted to specific needs for the different situations of use.

Very early in the project, we started to investigate into user needs and requirements, e.g. by interviewing visitors of a trade fair and by focus group discussions. Based on scenarios we discussed the system with prospective users and defined a set of essential use cases [5]. In several iterations, the user interfaces of desktop client and the mobile client have been improved, based on formative user evaluations of mock-ups. Visualization of maps, interaction with maps, and navigation support have been subject to extensive formative usability studies. Finally, we conducted a user evaluation of a working prototype at an international trade fair (Medica 2003 in Düsseldorf, Germany).

The remainder of this paper is structured as follows: First, we outline the functionality and the user interface of the SAiMotion system. In the second and third section, we report in more detail on user requirements of map visualization and navigation support. Fourth, we briefly report on issues concerning tour planning and tour visualization. Finally, we report on the outcome of the final user evaluation at the trade fair.

2 UI Concepts for a Nomadic Support

Concerning the user interface we faced several challenges:

- Designing this functionality for the small screen and restricted input means of a PDA, e.g. map display and search/filter functions on the small screen, interaction by pen input, and rather avoiding soft keyboard inputs;
- Nomadic support, i.e. supply similar functionality on desktop and mobile client with a consistent interface while making best use of each platform.

In the scenario-based user requirements analysis, we found information retrieval, interactive maps, and personalized tour planning to be of crucial importance for a nomadic exhibition guide. Accordingly the systems interface provides three main views:

- Information retrieval, where information of exhibits, exhibitors, and events is displayed in a tabular view; it provides various ways browse, sort, filter, and search in the large underlying content base.
- Maps, on several levels of detail; maps that highlight the current user position, a certain goal, a complete tour, and other points of interest.
- Tour in calendar view, to allow planning with timetable and appointments in view while the system supervises the spatial distribution of appointments and suggests a path-optimized schedule.

A couple of additional novel services have been implemented and demonstrated to the users, such as a news ticker and a buddy finder, both on the mobile client.[1]

The desktop client allows a good overview over complex information, such as the tour in calendar view, the index, and the information tables. While planning the visit on the desktop, the user looks for information that is to her or his personal interest or taste. This interaction is used to learn the personal preferences of a user. The user can also plan the tour around the exhibit, by first selecting exhibits or events he wants to visit, getting an automatically create tour, making changes to this tour etc. In this way the data become personalized. We chose an offline implementation, i.e. the data of the exhibition can be downloaded before the visit, and used for planning at home and while travelling. The personalized data can then be loaded to a PDA to support the mobile situation of use. Synchronisation mechanisms should be foreseen, to update the exhibition data from a central service, and to update the personalized data between PC and PDA.

Designing the UI for the mobile client requires in the first place to reduce functionality to the essential support, and then to cut the views into digestible chunks. The concepts used on desktop and mobile client need to be equivalent to facilitate users' roaming between both platforms. In the mobile situation, the complete database should be available, only the retrieval and display means have to be reduced to basic functionality. Basic retrieval is sorting and filtering by name of objects (companies, products, events), filtering by a flat index, and also searching by free text input. The flat set of keywords should be equivalent to the simple-mode filtering index provided on the desktop, only that the hierarchy is lost. Another important filter is by space (e.g. halls), as mobile users might look for a list of who is around.

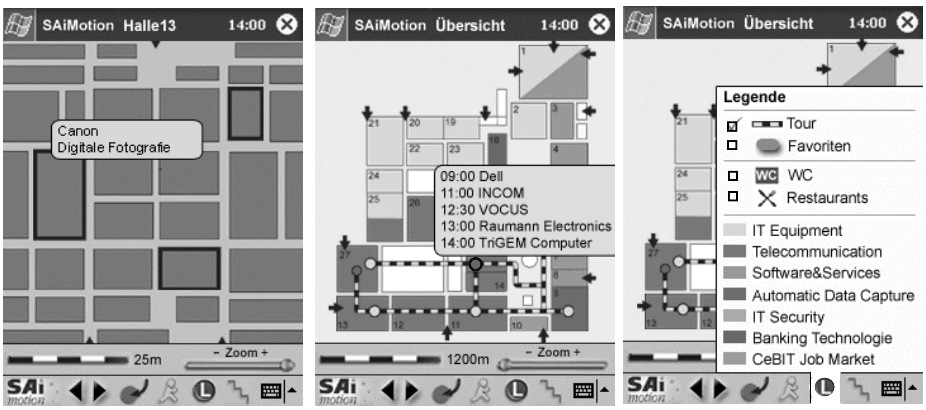
Concerning the amount of information given in one chunk, we decided to always show the complete list as resulting from a user request, which can be vertically scrolled. Only when sorted by name, there were separate lists for each letter in the alphabet. We also found that, even on the small screen, for each object in a list more than one line is needed, as users need not only the name of a company but some qualifying description as well. This will allow users to judge if they want to click on a name to see more details, and also to judge if the list contains what they were asking for (compare a similar finding in a user evaluation of a tourist guide [6]).

[1] These latter functions require wireless connection to server-based information services.

3 Map Visualization for Small Screen

To investigate user requirements for trade fair guides, several methods have been applied early in the project, including focus group technique and interviews of visitors to a fair [7]. The results showed a general need for the visualization of spatial information and for navigation support. Therefore, interactive maps were specified, including issues of appropriate scales, representation of tours, highlighting of topical selection of exhibits, and navigation support while going on a tour.

On the basis of this specification, a usability engineering process was implemented in the system development to ensure a sufficient ergonomic quality of the system. To test the interface conception for the interactive maps, an initial mock-up prototype was developed. It was based on HTML and comprised very limited functionality. However, it allowed to test most interaction sequences proposed in the interface conception: for critical and interesting interaction sequences, screen flows showing example content were realised and linked. The mock-up prototype ran on a web-pad with stylus-interaction. The screen size was assimilated to that of a PDA (Compaq iPaq).



(a) (b) (c)

Fig. 1. Prototype views of SAiMotion interactive maps, showing hidden labels in tooltips (a), tour elements and route together with additional information in a hidden label (b) and an interactive legend (c).

In a first iteration, a laboratory test with 15 users was conducted. The participants were introduced in the general usage of the HTML-prototype. The participants had to solve several tasks focussing on navigating with different map views, changing displayed information, manipulating map objects, especially elements of a tour, reading and integrating information from list- and map-views. To check the self-expressiveness of wording and iconic symbols, subjective and qualitative measures were recorded using the thinking-aloud method and semi-standardized interviews.

In a second iteration, the results of the first test were used to change the interface specification and to develop a second HTML-prototype. This prototype was also

tested with 11 users. Again, they were introduced to the basic interaction with the web-pad, and had to solve tasks.

The results of the second iteration cycle again were used to adapt the interface specification that finally became the basis for the graphical user interface for SAiMotion.

Abstract and Simplified Visualization. For small screens, complex spatial structures like a tour path must be simplified and generalized, especially in small scales. For instance, it should show only the sequence of halls but not the detailed route within halls (see figure 1 b). This type of map generalisation was easily understood by the participants of the user testing. However, the users wished to have the direction of the tour visualized especially in the large scales where the start and end points were not visible. Another important requirement was the use of colour coding of the exhibition map. This restricted the colour design heavily, especially the one of the fair overview. A problem of the map of the chosen German fair was that some colours were too similar to each other for PDA colour displays, so that the users could not distinguish between these colours.

Interactive Manipulation of Map Objects. The user should be able not only to display but also to browse and manipulate spatial data. E.g., a user should be able to select an object like an exhibitor on the fair from a list, to get a map on which the location of this object is highlighted, and to assign new attributes to this objects like being element of the user's tour or a point of interest. In our prototypes, this was realized by a context menu, which offered actions on particular objects, like adding an exhibitor to a tour. Especially the linking between map objects, texts, and listings was highly appreciated by the users in both tests.

Interactive and Dynamic Legend. The first prototype distinguished between two different types of legends [7]. The first one explained the coding of information, e.g. colours or symbols used for particular categories of map objects. This legend was adapted to the information currently displayed on the map view. The second type of legend was interactive, i.e. it allowed the user to add particular layers, e.g. to display a planned tour on the map. This interactive legend always showed the same set of items that were considered important attributes that should be directly accessible on any map view. However, in the first usability test the difference between those two types of legends was not clear to the users; they searched in both legends in order to read out codes as well as for changing displayed objects. Therefore, in the second prototype these two types of legend items were integrated in one legend, including stable items that can be ticked interactively, as well as temporary items that explain what currently can be seen on the map (see Figure 1 b).

Tooltips as Map Labels. To avoid cluttered screens, map labels can be hidden in tooltips that pop up when a map object is selected by clicking on it (see figure 2 a, b). This leads to a simple map view appropriate for small screens. The results of our two usability tests show, that tooltips are a very easy-to-use interaction style, which was immediately adopted by the users. They could retrieve object labels without leaving the small-scaled overview. The results also indicate that hidden labels for map objects are sufficient for *some* tasks. Users preferred not to leave the map view and to read additional information in the tooltips to build up a mental model of the spatial and

temporal structure of the tour. For other tasks, the majority of subjects did not use tooltips: When the users had to search for a known label, the tooltips where quite inconvenient to use. Therefore, hidden labels avoiding cluttered screens must be augmented by interactive elements like e.g. a list of map objects that allows selecting objects or a text-based to find map objects. Furthermore, important landmarks should be labelled directly on the map.

4 Egocentric Maps for Navigation

Egocentric maps adapt to the current position and orientation of the user: What the user sees in front of him in the building always corresponds to the map objects in the upper part of the PDA display above an ego-point. This is the typical design solution used, for example, in car navigation systems. Our hypothesis was that this type of maps also shows advantages compared to north-oriented maps for pedestrian navigation suitable for the SAiMotion fair guide. We expected, that the users can read the map easier and therefore produce fewer navigation errors such as wrong way decisions and are able to approach navigation goals faster. To test this hypothesis, a wizard-of-oz-study [8] was conducted comparing different design versions using a non-functional prototype that was controlled by a test moderator. Using simulated adaptive egocentric maps on a PDA, participants had to walk as fast as possible through an unknown and unclear area. This was compared to the usage of north-oriented maps that displayed the current position of the user and adapted the pan area but not the orientation of the view. In order to demonstrate expected effects of egocentric maps we chose a controlled experiment with 30 users randomly allocated to three conditions and investigated the differences using inference-statistical analyses. As expected, the egocentric view of maps supported the navigation best. The most impressive result was that test participants made absolutely no navigation errors when egocentric maps were used, in contrast to north-oriented maps. Also, they completed the navigation tasks fastest. Furthermore, the results also show that participants who were forced to produce an egocentric way by rotating the device manually also achieved good results in navigation efficiency, nearly comparable to automatically adapted egocentric views. However, an important limitation of this result is that in the test environment the alignment of the map was quite easy to do, because at each situation that required a re-alignment there were only few possibilities to match the displayed route to the corridors. It can be assumed that in less structured environments the manual rotation of north-oriented maps would be much more difficult and prone to errors.

We conclude that egocentric maps are a useful feature of mobile navigation devices, especially if manual rotation is difficult. If an electronic compass for automatic adaptation of the map alignment is available, this appears to be the preferable design solution. However, egocentric maps were not implemented in the SAiMotion demonstrators, because an electronic compass was not available in the final systems.

5 Tour Planning

Tour planning could be simple and straightforward: the user selects what he or she wants to see, and the system suggests a path around the exhibition site that covers all while optimizing the path. This is already a helpful service, which can be offered on desktop and mobile client. The requirements expressed by users, however, reveal more complex goals, varying interests, individual differences, and some computational issues. For example,

- Users may select too many points of interest to reasonably accommodate them in the time of their visit.
- Users have fixed appointments and the tour needs to include them without moving them. Some users wish a synchronisation with their familiar calendar tool.
- Users want to choose the duration of each visit according to their preferences and priorities.
- Users may want slack time for free ambling or recreation.

The planning algorithm can be made more sophisticated, but some user preferences cannot be inferred from existing knowledge about this user. This would leave the user with extensive task of declaring preferences, which is awkward and time-consuming. After all, the tour planning needs a clever algorithm plus user control and interaction. This requires a good visualization of the tour planner, as is feasible on the desktop but not on the small screen mobile client. The mobile client still allows creating a tour or making changes to an existing tour, but user control and interaction are restricted by the much simpler visualization. This seems adequate, as users, once on the exhibition grounds, lack time and patience for complex planning.

While testing with mock-ups we met two user views of a tour that are rather divergent: one user saw the tour mainly from his calendar point of view, the other mainly from the map representation. The user with the calendar view was a young manager of a company. His main interest in visiting the trade fair was to meet represent ants of companies. He would make appointments well ahead of his visit and leave only little time in his agenda to fill with seeing more exhibits that happen to be on the way. The other person was a scientist who would try to tour the grounds looking for interesting novel exhibits. She had a mainly spatial view on the tour, and wanted support to optimize the path, not to miss any object that might match her interests. Our conclusion is that both views must be provided, and users must be able to switch easily between the calendar view and the map view of a tour.

6 User Evaluation

A prototype of SAiMotion was implemented and demonstrated at the Medica ‘03 in Düsseldorf, in November 2003. The content covered all exhibitors, i.e. some 4500 companies, but exhibits only as far as companies had provided these data. In one hall,

localisation was implemented by means of WLAN positioning. The WLAN was not used for data transmission but for positioning only (in fact, it was other companies' infrastructure). A simple version of the interactive maps discussed above was implemented, i.e. a map of the whole site and a detailed map of this hall, also indicating the user position and the tour. The mobile client could be used in the hall for navigation and also to look up information by searching and filtering in the whole fair content. The desktop client was shown and could be tried on the fair, but had not been made available for visitors prior to the fair. This setting did not provide the full personal experience of using SAiMotion as intended, but was sufficient to give the users a realistic impression in the real environment. The desktop had previously been subject to an evaluation by 4 users in the usability lab at Fraunhofer FIT.

One desktop client and two mobile clients were available. Visitors were offered the opportunity to try the system and then asked to fill in a questionnaire. Within 3 days, 38 visitors took the opportunity to have a closer look at the prototype and also filled in the detailed questionnaire. Some more visitors saw the system but did not fill in the questionnaire, mostly because of lack of time. The questionnaire asked for

- A few demographic data, such as age and gender;
- Users' familiarity with computers and WWW;
- Their habits concerning trade fairs, such as how many visits per year, and information needs for preparing and while visiting;
- Assessment of the SAiMotion system, both for desktop and mobile;
- A ranking of features, both for desktop and mobile client;
- Overall evaluation of the benefits of a nomadic exhibition guide.

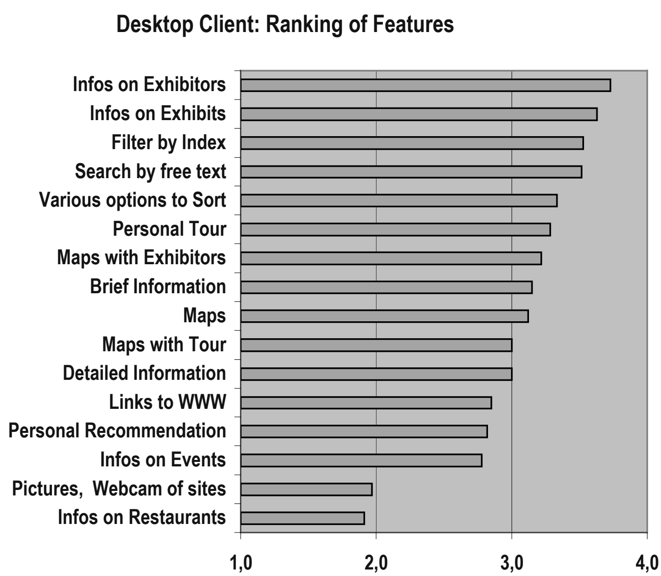


Fig. 2. Ranking features of a desktop client (on a scale of 1=unimportant to 4=very important)

Overall, a majority of 80% of users considered the system a valuable support for preparing and visiting trade fairs, compared to information and media available so far. A narrow majority of 54% said they would also be willing to pay an extra for using the system. The amount of extra they considered appropriate ranged between 5 and 10 EUR for one exhibition. Independent of the implementation in SAiMotion, the results of the ranking of features are valuable for future realizations of exhibition guides.

It is also obvious from these results that the situation of preparing the visit on the desktop and the mobile situation during the visit itself require different support. Figure 2 shows diverging importance of features, where users distinguish the mobile and the desktop usage. The mobile is not a downscaled web portal but a specialized support to fit the mobile situation. The challenge is to design both interfaces for a homogeneous look and feel, as far as possible.

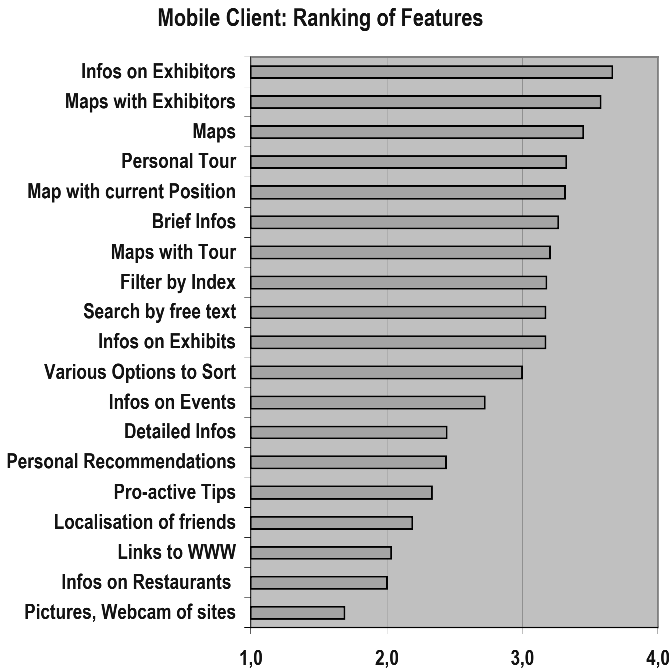


Fig. 3. Ranking features of a mobile client (on a scale of 1=unimportant to 4=very important)

The results also suggest that the users rank basic support such as information retrieval much higher than sophisticated features such as buddy finding or news ticker. We suggest that the high-ranking features are crucial for a widespread acceptance and need to be of best possible quality in order to justify the users' extra expenses. Whether the lower ranking features would be seen as attractive added values is a matter that needs more investigation. Our hypothesis is that for a business person with a limited time budget, the core features are just about all they want. They are not looking for extra fun and entertainment but for means to filter the information over-

flow that is already typical for a trade fair. This may, however, be different for people who visit a trade show for private reasons, or for other types of exhibitions where personal experiences and fun are main reasons for a visit.

Another indicator of users' request for continuous support from the planning stage up to the post-visit stage, was a request we had as an early requirement and also in the final evaluation: users asked for a feature that allows them to annotate information. They would like to make some notes in preparation of an appointment with an exhibitor, and also to annotate information about an exhibit when they have seen it, for future reference. They intend to collect information including their own notes, to be used and evaluated after their visit.

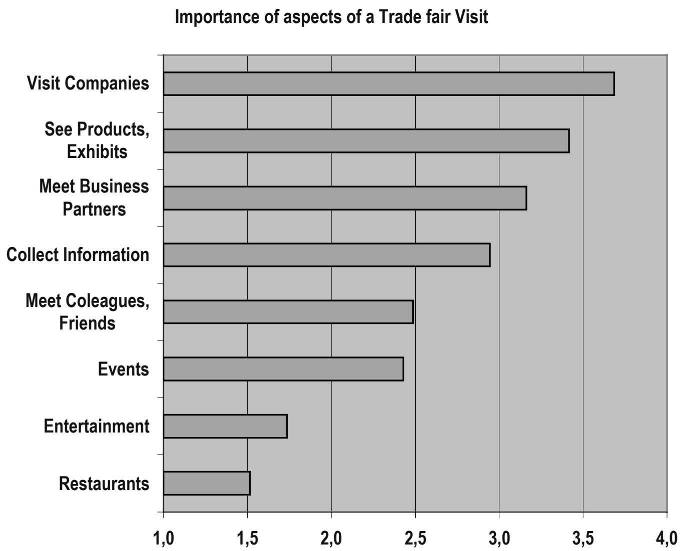


Fig. 4. Importance of aspects of a visit to a trade fair (on a scale of 4=very important to 1=unimportant)

The user assessment concerning pro-active recommendations, news-ticker, web cam pictures, or buddy finder has not been enthusiastic. The average rank of all these services was below 2.5 on our scale, see figure 2. One explanation might be that in the overwhelming information noise of a trade fair, all additional information services that are not strictly to a user's most urgent information needs can easily become a nuisance instead of help. Also, trade fair visitors ranked the aspect of meeting friends on the fair rather low, together with other aspects of entertainment and fun, see Fig. 3. The latter may be quite different, however, on a fair that explicitly addresses private people, consumption and entertainment. But also in this case, the local arrangements address users' attention and senses in a way that leaves no room for "playing around" with a mobile guide that tries to provide additional entertainment. Related to proactive recommendations, the organizers voiced a concern that the exhibitors might to be suspicious of such a tool they want to be treated fair and equal, they might not trust a

recommender tool that ranks them, rendering information to the user in a "biased" way.

7 Conclusions

Overall, the concepts of SAiMotion have been confirmed. Especially the nomadicity axiom has been corroborated, i.e. that mobile situations are only part of a greater context that needs continuous and consistent support. The mobile situation has its own requirements, as have planning of the visit and evaluation after the visit. For the functionality of the mobile and information presentation, it seems crucial to design for simplicity, i.e. to limit the functionality to the essential use cases and to strip the layout from all superfluous and fancy extras. Switching easily between map view and list view of retrieved information and the tour in calendar view is essential, and must be easy and natural for the user.

Maps are among the key features of the mobile support, and in spite of limited screen size can be rendered in a sufficient quality. Highlighting the user position, the tour, or a certain goal make maps even more valuable. The precision of WLAN tracking is still varying and not quite reliable. One can hope for improved positioning technology to come soon. Another challenging track of research might be to investigate how to make best use of unreliable positioning information.

Though our first approach was already well-considered and based on expertise in the field of mobile guides, it has paid off to invest in a user-oriented engineering process in order to make the appropriate decisions.

Acknowledgements. This research has been conducted in the context of the SAiMotion project, which was a joint project of several Fraunhofer Institutes, i.e. FIT (coordinator), IAO, IGD, IIS, IPSI, and IZM. We wish to acknowledge our partners' contributions. SAiMotion has been funded by the "Bundesministerium für Bildung und Forschung" (BMBF, Funding-Key 01AK900A). The authors are responsible for the contents of this publication.

References

1. Abowd, G.D., C.G. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton: Cyberguide: A mobile context-aware tour guide. Wireless Networks, 1997. **3**: p. 421-433.
2. Cheverst, K., N. Davies, K. Mitchell, A. Friday, and C. Efstratiou. Developing a Context-aware Electronic Tourist Guide: Some Issues and Experiences. In CHI 2000. 2000. Netherlands.
3. Oppermann, R. and M. Specht. A Context-sensitive Nomadic Information System as an Exhibition Guide. In Handheld and Ubiquitous Computing Second International Symposium. 2000.
4. Schmidt-Belz, B., H. Laamanen, S. Poslad, and A. Zipf. Location-based Mobile Tourist Services - First User Experiences. In ENTER 2003. 2003. Helsinki, Finland: Springer.

5. Constantine, L. and L. Lockwood, Software for Use. ACM Press. 1999, New York: Addison-Wesley.
6. Schmidt-Belz, B. and S. Poslad. User Validation of a mobile Tourism Service. In HCI in mobile Guides. 2003. Udine, Italy: University of Udine.
7. Hermann, F. and F. Heidmann. User Requirement Analysis and Interface Conception for a Mobile, Location-Based Fair Guide. in 4th International Symposium Mobile HCI. 2002. Pisa, Italy: Springer, Berlin.
8. Hermann, F., G. Bieber, and A. Duesterhoeft. Egocentric Maps on Mobile Devices. In 4th International Workshop on Mobile Computing, IMC 2003. 2003. Rostock: IRB Verlag, Stuttgart.

Social Networks and Mobile Games: The Use of Bluetooth for a Multiplayer Card Game

Chris Baber[1] and Oliver Westmancott[2]

[1] Electronic, Electrical & Computer Engineering, The University of Birmingham, Edgbaston, Birmingham. B15 2TT UK
[2] Westmancott Consulting, The Big Peg, 120 Vyse Street Birmingham. B18 6NF UK

Abstract. In this paper, we consider mobile game playing from the perspective of social network analysis. A multiplayer card game (BELKA) has been designed. The game allows players to select between trading, playing or pairing with other players. The game was played using playing-cards or using HP Ipaq devices equipped with Bluetooth, and players were either seated around a table or encouraged to move around the room. Activities during play were recorded and these data are analysed in terms of social networks. It was found that while the playing-cards led to attempts to apply the same type of activity in both seated and mobile conditions, the use of PDA led to differences in play. These differences were due to both technical, i.e., availability of players in the Bluetooth network, and social, i.e., visibility of players in the world and the activity of the Dealer. It is proposed that the manner in which the game was played changes when a mobile device is used whilst moving around, and that this is different to when the same device is used when sitting down.

1 Introduction

Mobile games can be said to require players having equipment that can be carried in their pockets and that allows people to set up and play a game wherever they are. A deck of playing-cards obviously fits these requirements and the number of games that have been developed to make use of these artifacts are legion [Hoyle and Dawson, 1994]. Taking the notion that playing-cards represent the *de facto* standard for mobile games, this paper considers how playing a card-game on a Personal Digital Assistant (PDA) differs from using actual playing-cards. To explore the issue of ‘mobility’ a game was designed to encourage social interaction and an experiment devised that allowed a balanced comparison of mobility (sitting or moving) and media (cards or PDA).

In terms of what might be termed the social aspects of mobile gaming, mobile devices typically support game playing between *pairs of individuals*. Devices can be connected on a one-to-one basis, via cables, to allow people to play some games in a two-person mode. Bluetooth (and infra-red) have been used to remove the cable, but the game play is still typically one-to-one. In the research reported in this paper, we were interested in exploring games in which multiple players could participate simultaneously and which could be played wirelessly, using Bluetooth. Of course, Bluetooth is usually employed as a peer-to-peer communication channel and this

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 98–107, 2004.

might not be ideal for multiplayer gaming (use of a wireless local area network and server could have supported more efficient communications), but it was felt that this represents a contemporary wireless communication medium that can support game playing anywhere. Thus, one might envisage two people on a bus, both carrying handheld devices being able to search, via Bluetooth, for other players and challenge them to a game, without the need to log on to a server.

1.1 (Technologically-Mediated) Mobile Gaming

One definition of a 'mobile game' is simply a computer game that can be played on a portable platform. Thus, the device can be used more or less anywhere (either because it is self-contained or because it can access communications services, like a mobile telephone). However, this only conveys some of the attributes of mobility that are related to gaming. For example, a chase game could involve movement of players (in real or virtual space), and awareness of the activity of other players; a challenge game might require access to particular services and access to other players'. Table 1 uses this extended notion of mobility to contrast mobile gaming studies. The games considered in table 1 are either chase, in which players need to capture other players (the 'other players' might be represented in 'virtual' space, e.g., on a website), or challenge, in which players perform specific actions in order to win the game. The supporting functions are provided either by the technology used (T), through social interaction in the real world (S) or through activity in a virtual space (V). For this study, play occurred in a 'real' (as opposed to virtual) space, with players free to move around, and involved the exchange of 'virtual' playing cards through wireless connection. There is no location-awareness for this game, although it is proposed that in terms of the classification proposed in table 1 the game is still clearly 'mobile'.

Table 1. Comparison of games in terms of attributes of mobility

Game Type	Mobile	Aware: Location	Aware: Player	Access: players	Access: services	Move: play	Move: access	Reference
Chase	T				T	S		Rheingold, 2003
Chase	T			V	T			Longmate and Baber, 2004
Chase	T	T	V	V	S	SV		Flintham et al., 2003
Challenge	T	S	S	T		TV		Bjork et al., 2003
Challenge	T		SV	ST	V	S	T	This paper

1.2 Games as Social Activities

It is ironic that one of the main reasons that people play games (for social contact and interaction) is not easily supported by mobile technology. In this paper we will consider the social aspects of game playing, using a form of social network analysis [Wasserman and Faust, 1994]. In order to further explore the social dimensions of

game playing, this project will consider the ways in which moves are made in the game. By move, we mean a sequence of actions that make up a complete set. For example,

Player A challenges Player B to {play, trade, pair}
Player B {accepts, rejects}
Player A offers a card (if Player B accepts a play or trade)
Player B offers a card
Player A or Player B collects the cards (if they have won a play)

We propose that a move can be seen as a series of actions, with each action being regulated by simple rules. It is probably important, at this point, to note that the game has been designed to encourage as many moves as possible to occur concurrently. In other words, any pair of players can perform a move as long as they are not currently involved in a move – this means that, in a game of 6 players and 1 dealer, it is possible to have 3 moves occurring at the same time.

2 BELKA

A popular class of games for mobile devices is card games. These games are, of course, single player games. However, we thought that a card game for multiple players would be appropriate for this project for the following reasons:

- simple graphics requirements;
- simple communications requirements;
- straightforward game-play;
- easily interpretable symbols, rules and conventions;
- easily integrated into social activity, such as talking to other players.

The game invented for this research, BELKA, is a combination of a trading game (in which players can swap cards), a challenge game (in which players attempt to win cards from other players by playing a higher status card), a collaboration game (in which players can pair-up to produce a better hand), and poker, in which a winning hand is defined using definitions from 5-card draw. Players begin the game with three cards each and, through playing or trading or pairing, seek to collect a 'good hand' from which they can select up to 5 cards to present for judging at the close of the game. As mentioned previously, playing, trading and pairing can be performed by any pair of players and it is intended that several pairs can play at any point in time. This means that turn-taking is not defined by any hard and fast rules (except that a player may only participate in one move at a time).

2.1 BELKA on HP IPAQ 5450

The basic elements of BELKA were programmed, using Visual Basic.Net onto six HP Ipaq 5450 handheld devices and a Toshiba Satellite 1730CDT laptop. Communication

is managed through Bluetooth; this was built-in to the Ipaqs and was put onto the laptop using a Belkin Bluetooth USB Adapter. The laptop assumes the role of the Dealer and distributes cards to each PDA. The Dealer also supports trading with any PDA and updates trumps (which it passes onto any PDA that connects to it). At the end of the game, all PDA upload their gamelogs into a directory on the laptop. The HP Ipaq 5450s are used as the devices on which players view their cards and make their plays. However, it is worth noting that these devices are essentially dumb as to the rules of the game. If the rules had been hard-coded into the devices, then much of the social activity that was of interest might have been subsumed into simple acts of button pressing.

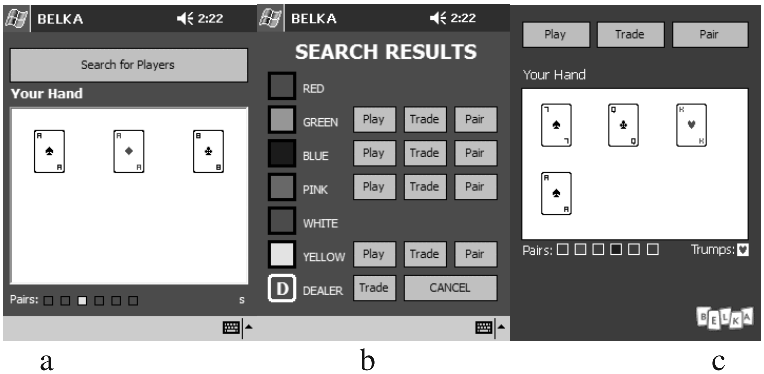


Fig. 1. Once players log onto the Dealer, then they receive three cards (a). On searching for players, they are shown a screen of available devices (b). At this stage, players are able to initiate plays (c)

Initial trials of the system revealed a potential shortcoming of Bluetooth - not all devices were visible via Bluetooth (even if the players were standing next to each other). However, this was exploited to advantage in the game in that play has to be negotiated both socially, in terms of challenging a person to play and moving to that person, and technically, in terms of seeing an available device and making a connection. When a player is available, then there are three options (play, trade, pair) that can be made. The option is selected and this sends a request between devices. The challenged player is then able to Accept or Reject the option. If the option is Accepted, then cards are selected and played (or exchanged). After the move, the initial screen is updated.

3 User Trial

A user trial was conducted with the equipment. The trial involved 38 people and lasted around 5½ hours. All participants were Undergraduate students on the MEng Interactive Systems course (mean age approximately 19 years; 14 female; 24 male). Participants were assigned to groups of 6 or 7. The study employed a repeated-

measures design. It was felt that the nature of the game was such that transfer effects between conditions would not hamper any findings.

3.1 Method

An initial introduction to the rules of the game and a 90 minute training session were used to ensure that all players had a reasonable understanding of the game and its rules, and also that they were comfortable playing the game using Cards or Ipaqs. In particular, we wanted to ensure that players would, where possible, call each other by their game colour (Red, Yellow, Purple, Blue, Pink, White) and to use a simple but formal challenge, i.e., <White> trades with <Pink>. The following week, the main trial took place. In the main trial, players were then assigned to three groups. Participants in a group were further divided into three teams of 6: one team to play using cards, one team to act as Observers for the card players and one team to play using the PDAs. Each team played two practice games and one formal game whilst seated and whilst moving under each condition. This resulted in a total of 12 games played by each team. The games were time limited (to 3 minutes each) and the Experimenter called Time at the end of this period and adjudged the winning hands.

Table 2 shows the design of the experiment. The experiment was designed so that the optimal settings for each condition were contrasted, e.g., the playing cards are best played sitting down at a table, the PDAs are best used when moving around. In order to develop a balanced experimental design, it was decided to contrast cards x PDA in terms moving x sitting. It is clear that playing cards whilst moving is a little odd, after all cards are intended to be played on a flat surface rather than hand-to-hand, so one might anticipate different performance between Cards + Sitting and Cards + Moving conditions. It is less clear as to why PDA + Sitting and PDA + Moving ought to lead to differences; the PDA requires the user to hold it in one hand and act upon it with the other whether they are standing up or sitting down, so the differences might not be so apparent. However, by allowing all variations, we should be able to contrast the effects of playing a card game when it is performed using real cards and when it is performed using their digital counterparts.

Table 2. Design of User Trial

	Sitting	Moving
Cards	A	B
PDA	C	D

4 Results

During the game, moves were recorded (either manually by Observers, or automatically by the PDAs). These data are used to describe overall performance, in terms of moves, and to describe the characteristics of the networks for each condition.

4.1 Overall Performance Metrics

Table 3 shows the average performance data, across 4 games for each condition, i.e., mobility x medium. The percentages indicate the relative contribution of the type of move to the total number of plays. Comparison of tables two and three indicate that the Card conditions led to far more moves than the PDA condition. The difference in number of moves between Card and Ipaq conditions was due to the constraints that the PDA and Bluetooth imposed on the interactions; players needed to establish a connection and then to exchange tokens over Bluetooth. Thus, the Card condition could perform almost twice as many moves than the PDA condition. This led to a more frenzied game in the Card conditions. Indeed, the Card conditions involved players crowding the Dealer in order to check when trumps would change and adopting a fairly conservative approach of trading with the Dealer as far as possible.

For the Cards + Moving condition, there are fewer moves in total when moving around, possibly due to the fact that exchange of cards had to be performed hand-to-hand rather than on a table, possibly because of the need to move from place to place to perform a move. However, the players in the Card + Moving also tended to crowd the Dealer which implies that they were playing the Moving game in very much the same manner that they played the Sitting game. The difference in number of rejected plays implies a change in tactic. When Sitting, players have a clearer idea of what plays have gone before and whether a particular player holds a strong card, but when moving this information is less easy to obtain.

The PDA conditions differ from the Cards conditions not only in terms of the number of moves, but also in the distribution of moves. The PDA conditions led to fewer interactions with the Dealer and more plays between players than the Cards condition. On the basis of these data, it would appear that the PDA conditions also led to similar performance across Sitting and Moving. Thus, one might assume that these mobile devices were used in much the same manner whether people moved or were stationary. However, observation and discussion with participants suggested that this was not the case and the following Social Network Analysis is presented to explore possible differences between Ipaq conditions.

Table 3. Comparison of mean performance data across conditions (Seated on left; Moving of right)

Activity	CARD	PDA		Activity	CARD	PDA
Turns	56.5 (100%)	21.75 (100%)		Turns	33.25 (100%)	18.25 (100%)
Trades (+ Dealer)	27.5 (48.7%)	10.25 (47%)		Trades (+ Dealer)	18.5 (55.6%)	6.75 (36.9%)
Trades (+Other players)	2.75 (4.8%)	1 (4.6%)		Trades (+Other players)	0.25 (0.8%)	0.5 (2.7%)
Rejected plays	9.25 (16%)	0.5 (2.3%)		Rejected plays	1.75 (5.3%)	0.25 (1.4%)
Pairs	1.25 (2.2%)	0.25 (1.2%)		Pairs	0.75 (2.3%)	0.5 (2.7%)
Plays	15.75 (28%)	9.75 (44.8%)		Plays	12 (36%)	10.25 (56%)

4.2 Social Network Analysis

Conventionally social network analysis takes the form of a binary link or no-link between people and then calculates the amount of network activity relative to the total possible amount. The calculations could indicate the centrality of a given individual or the number of people who need to be connected in order to support a particular link. In this study we were interested in the amount of interaction between players. Thus, a simple binary distinction would be inappropriate. In this analysis, we present the mean number of interactions per game for each condition (averaged over the course of 4 games for each condition). In order to calculate the degree of contribution to a game, we relate this figure to a total number of moves for a given game. From this, it is possible to determine whether a given player has made a large or small contribution to the game. The figures in the cells also allow us to indicate any interactions between pairs of players that seem particularly high[1]. The data from the studies are shown in tables in the Appendix.

CARDS + SEATED: The seating plan for the Cards + Seated condition is shown in figure 2. From table A1, we can see that the Dealer (D) has around 60% of the moves (which is born out by the observation that many of the moves in these games involved players trading with D). From table A1, the highest proportion of plays involve the Dealer, or are between Blue and Green or Purple and Green (who are facing each other) or between Blue and Purple (who sat next to each other).

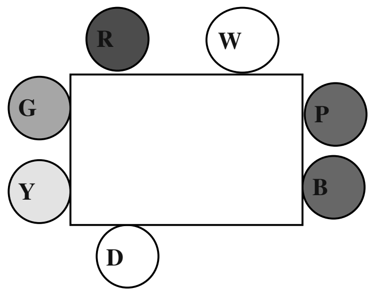


Fig. 2. Seating plan for Cards + Sitting

This observation implies that:

a. the Dealer dominates the game;
b. a few players tend to dominate the game, perhaps with other player being sidelined by these four (D, B, G, P);
c. that moves tend to follow a line-of-sight, i.e., opposite to each other on the table or next to each other.

[1] The data in the tables are presented in terms of Colours, rather than people. Each colour indicates a role that was assumed by players in a game. As we are averaging across 4 games, this means that each set of data is an average of 4 people's performance. For the sitting condition, we feel that this is justified in that the position of the colours was kept constant across games. Thus, the Dealer always sat in a specific chair and the players were always positioned in the same seat relative to each colour. For the moving condition, the averaging of data by Colour is less easy to defend; players were free to move around and there was little if any fixed position for them. However, we feel that averaging by colour provides a convenient basis for comparing sitting and moving and gives some insight into the way in which the groups of players interacted in each condition.

PDA + SEATED: The seating plan for the PDA + Seated condition is shown in figure 3. Table A2 shows that the Dealer (D) has around 50% of the moves, as with the Cards condition. Also, if one explores the moves between player in table A2, then one can see that Red-Yellow, Pink-Yellow and Green-Yellow have a higher proportion of moves than other combinations. As with the Cards + Seated condition, line-of-sight is a strong indicator of whether two players will join on a move.

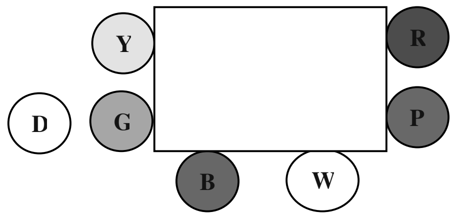


Fig. 3. Seating plan for PDA + Sitting

CARDS + MOVING: In the Cards + Moving condition (see table A3), over 60% of moves involve the Dealer. Observing this game, and speaking to players afterwards, it was clear that playing this game using Cards whilst moving was treated in much the same manner as the Cards + Seated condition. Most players as near to the Dealer as possible, which allowed them to check changes in trumps and to trade with the Dealer. Plays between players were made whilst moving around the Dealer. The lower number of rejected plays implies a lack of knowledge about who were the stronger players.

PDA + MOVING: Dominance by the Dealer is less marked in this condition, and (with the exception of Yellow), all players seem to make a similar contribution to the game (see table A4). After the game, players confirmed that their choice of moves was defined by who was available, i.e., in terms of Bluetooth connectivity. A move involved players physically walking over to the player they were challenging. It would seem that the mobile devices, therefore, led to far more movement than the playing-cards.

5 Discussion

This study has demonstrated that some aspects of mobility have a significant impact on play. It is interesting to note that, when playing on a mobile device (whilst seated), people would try to adopt a similar strategy and game-play to that used for playing-cards. It was only when the players used the mobile device *and* moved around that more marked differences between Cards and PDA were apparent. This study has demonstrated that mobile gaming is not simply a matter of playing games on mobile devices, but that the essential ability to support moving around changes the nature of play and alters the social aspects of gaming. Furthermore, the role of Dealer clearly has a significant bearing on play. In the Card conditions, the Dealer is an active and significant member of the group and is involved in more moves than the other plays. In the PDA condition, the Dealer is less active and, particularly in the mobile condition, is not involved in so many moves.

Through the use of a controlled experimental design, the study suggests that mobility is both a feature of and a consequence of playing a multiplayer game through Bluetooth. The instability of the connection offered by Bluetooth could be exploited as a means of supporting interaction between players in a game, e.g., by forcing strategic adaptation to the availability of other players. What is, perhaps, of more interest in the observation that playing the game on PDA alone (i.e., with no movement) led players to adopt strategies that we very similar to playing the game using cards, and that playing with cards whilst mobile also caused players to retain their strategies from the seated condition. The implication is that developing mobile games is not simply about placing conventional play onto a handheld platform, but requires consideration of the interplay between the technical, virtual and social aspects that were considered in table 1.

Acknowledgements. This work was supported by a Microsoft Embedded Systems Research grant (APM-UK-01b)

References

1. Hoyle, E. And Dawson, L.H. (1994) *The Complete Hoyle's Games*, London: Wordsworth Editions Ltd.
2. Rheingold, H. (2003) *Smart Mobs: the next social revolution*, New York: Perseus
3. Longmate and Baber (2004) Integrated digital communities: combining web-based interaction with text messaging to develop a system for encouraging group communication and competition, *Interacting with Computers*
4. Flintham, M., Anastasi, R., Benford, S., Hemmings, T., Crabtree, A., Greenhalgh, C., Rodden, T., Tandavanitj, N., Adams, M. and Row-Farr, J., 2003, Where on-line meets on-the-streets: experiences with mobile mixed reality games, *CHI'03,* New York: ACM
5. Björk, S., Falk, J., Hansson, R. and Ljungstrand, P., 2001, Pirates! Using the physical world as a game board, In M. Hirose (ed), *Interact01*, Amsterdam: IOS Press, 423-430
6. Wasserman, S. & Faust, K. (1994) *Social Network Analysis: methods and applications*, Cambridge: Cambridge University Press

Appendix

Table A1. Interactions between players in the Card: Sitting condition

	Dealer	Red	Green	White	Purple	Blue	Yellow
Dealer		4.75	4.25	4	4.5	4.75	5
Red	4.75		1	0.75	0.75	1.25	1.5
Green	4.25	1		1	3.25	3	0.75
White	4	0.75	1		0.5	0.5	0.5
Purple	4.5	0.75	3.25	0.5		3	1.25
Blue	4.75	1.25	3	0.5	3		2
Yellow	5	1.5	0.75	0.5	1.25	2	
Degree	0.57	0.21	0.28	0.15	0.28	0.3	0.23

Table A2. Interactions between players in the Ipaq: Sitting condition

	Dealer	Red	Green	White	Purple	Blue	Yellow
Dealer		1.5	1.75	1	1.75	2	2.25
Red	1.5		0.75	0.25	0.75	0.25	1.75
Green	1.75	0.75		0.25	0.5	0.75	1
White	1	0.25	0.25		0.75	0	0.5
Purple	1.75	0.75	0.5	0.75		0.75	1.5
Blue	2	0.25	0.75	0	0.25		0
Yellow	2.25	1.75	1	0.5	1.5	0	
Degree	0.54	0.28	0.27	0.14	0.28	0.19	0.36

Table A3. Interactions between players in the Card: Moving condition

	Dealer	Red	Green	White	Purple	Blue	Yellow
Dealer		2.75	1	1	6	6.75	1
Red	2.75		0.5	1.5	1.25	3.25	0
Green	1	0.5		0	0.5	1.75	0
White	1	1.5	0		0.75	0	0
Purple	6	1.25	0.5	0.75		2.25	0.25
Blue	6.75	3.25	0	0	2.25		0
Yellow	1	0	0	0	0.25	0	
Degree	0.64	0.32	0.07	0.11	0.38	0.49	0.04

Table A4. Interactions between players in the Ipaq: Moving condition

	Dealer	Red	Green	White	Purple	Blue	Yellow
Dealer		1.75	1	1	1	1	1
Red	1.75		1.5	0.75	0.5	1	0.5
Green	1	1.5		1.25	1.75	0.75	
White	1	0.75	1.25		0.25	0.5	0.5
Purple	1	0.5	1.75	0.25		0.5	0.25
Blue	1	1	0.75	0.5	0.5		0.25
Yellow	1	0.5		0.5	0.25	0.25	
Degree	0.39	0.36	0.37	0.25	0.25	0.24	0.15

Eye Movement Study of Reading on a Mobile Device Using the Page and RSVP Text Presentation Formats

Gustav Öquist[1], Anna Sågvall Hein[1], Jan Ygge[2], and Mikael Goldstein[3]

[1] Uppsala University, Department of Linguistics and Philology
Box 635, SE-751 26 Uppsala, Sweden
gustav@stp.ling.uu.se, anna.sagvall_hein@lingfil.uu.se
[2] Karolinska Institutet, Department of Clinical Neurosciences
Polhemsgatan 50, SE-112 82 Stockholm, Sweden
jan.ygge@ste.ki.se

Abstract. We present findings from a balanced repeated-measurement evaluation where 16 subjects read texts of similar length and difficulty using the traditional Page and the dynamic Rapid Serial Visual Presentation (RSVP) format on a mobile device. Apart from monitoring reading speed, comprehension, and NASA-TLX task load, we also devised a system that enabled us to keep track of subjects eye movements. The results indicate no significant differences in reading speed or comprehension, but for task load, RSVP increased the Temporal task load factor. However, the most striking differences were found in the eye movement recordings. RSVP was found to decrease the overall number of eye movements significantly. But, RSVP was also found to significantly increase the number of regressions, although it decreased the number of saccades. These findings contradict common claims and their implications for the improvement of readability on mobile devices are discussed.

1 Introduction

The catch with readability on mobile devices is typically that mobile devices have to be small to be mobile, whereas the traditional page format used on a limited screen space is at odds with how we are accustomed to read. So, what can we do to resolve the issue? In short, there are two options. One is to enlarge the screen, the other is to change the way we present information. From a technical perspective, both options are equally viable, but from a consumer perspective, enlarging the screen is unacceptable as this implies a device with less portability [20]. Thus, it seems that we need to change the way we present information on small screens to achieve better readability. Yet, is this *really* a viable option just because enlarging the screen is not? To learn more about this, we have chosen to study eye movements when reading on a mobile device using two equally efficient, but very different, text presentation formats.

[3] Valhallavägen 130, SE-114 41 Stockholm, Sweden. mikael.goldstein@telia.com

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 108–119, 2004.

2 From Eye to Readability

In order to be able to read, the image of the text must be projected upon the *retina*, a thin membrane with photosensitive receptors located on the back of the eyeball. Although the retina has a 240-degree field of vision, the maximum resolution is restricted to the fovea, which the area where the fixation target must be located for accurate recognition [17]. The foveal field of vision is only one or two degrees wide and this means that only six to eight characters can be in focus at a time. The parafoveal region immediately surrounding the fovea further extends the perceptual span to approximately 20 characters, but beyond that acuity is too low for retrieval [13]. The consequence of this for reading is that we have to move a very narrow focal point of vision across the text to be able to read it. Reading a text with a spatial layout, like on this page, consists of three distinct visual tasks: processing information in fixed gazes or *fixations*, performing *saccadic* eye movements to move between fixations, and moving to the next line using *return sweeps* [15]. The length of a saccade stretches between 1-20 characters. A saccade must not be directed forward in the text. Backward saccades, or *regressions*, are executed for going back and reread text. When reading on paper, about every fifth saccade is a regression [6].

Based on the observations of eye movements while reading, we may say that reading is a process of determining when to move to the next fixation target and where to move the eyes next. One of the most influential models of reading is Just and Carpenter's Reader model [6]. It was based on empirical eye movement data and stipulates the following two assumptions: The *immediacy hypothesi*s, which states that each word is immediately processed when it is fixated, and the *eye-mind hypothesi*s, which states that the eyes remain fixated on a word as long as it is being processed [6]. Both assumptions have later been criticized because they don't account for context and parafoveal preview effects [11, 16]. Yet, if we combine Just and Carpenter's processing model with oculo-motor modeling of the physiological limits of the eye and the visual properties of the text, we may get closer to a realistic definition. Fixation duration, i.e. determination of when, is assumed to be governed by cognitive processing, while saccade execution, i.e. determination of where, is governed by a combination of linguistic, orthographic and oculo-motor factors.

Readability is typically referred to as the ease with "which the meaning of text can be comprehended" [7:331]. Readability is often measured in terms of *reading speed* and *comprehension* based on actual reader performance [7]. Reading speed is calculated as words read per minute (wpm) whereas comprehension is represented as percent of correctly answered questions posed on the text. Measuring readability by reading speed and comprehension yields an objective approximation of reading efficiency. In order to learn more about the subjective experience, additional attitude or task-load inventories have to be considered as well. What we want to learn by measuring readability is to find the text presentation format that best support reading. Given that the goal of reading is to comprehend a text, a natural assumption would be that a high readability score should reflect that the reading process has proceeded efficiently. An alternate definition of readability that relates closer to how we actually read may thus be: the ease with which the reading process can proceed.

3 Reading on Mobile Devices

It is evident that text presentation on small screens does not work well enough to be satisfactory today [2]. One could argue that readability on mobile devices is likely to improve with increased resolution as it did on desktop screens [9]. However, the problem with readability on small screens is not so much the resolution as the physical limitation in the screen space. We are used to read text presented in the page format, which demands far more space than most mobile devices can offer. There exist a few different solutions for text presentation on small screens that employ different techniques to handle the problem [7, 8]. They can be divided into *traditional* and *dynamic* text presentation formats. The major difference being that traditional text presentation preserves the spatial layout of the text whereas the dynamic trades a spatial dimension for a temporal and presents the text over time.

3.1 Traditional Text Presentation

The benefit of *traditional text presentation* is that the formats are familiar to the readers as text is presented in the same manner as on a paper page. However, since a full page cannot be displayed on a small screen it is divided into smaller parts. The text can then be presented either as smaller pages that fit the screen, a technique called *paging*, or as a long page continuing outside the screen, a technique called *scrolling*. In the page format turn-page keys are used to move between the pages and in the scroll format a scroll-bar is used to move in the text. Continuous scrolling has been found to be preferred compared to step-by-step scrolling, but the Page format is more liked for reading [7]. Most programs dedicated to reading, for example e-book readers, make use of traditional text presentation in the Page format. *Microsoft Reader* is an example of an e-book reader that is available for mobile devices such as Personal Digital Assistants (PDAs) (Fig. 1 left). The interface is designed to remain as similar to a real book as possible and close attention has been paid to the traditions of proper typography, it moreover utilizes ClearType to enhance legibility [5].

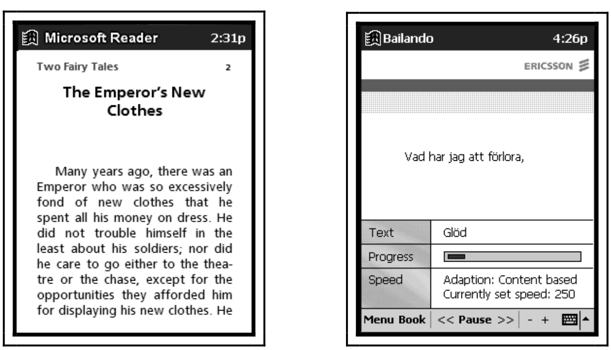


Fig. 1. Microsoft Reader, an interface using the traditional Page format (left), and Bailando, an interface using the dynamic RSVP format (right)

3.2 Dynamic Text Presentation

The benefit of *dynamic text presentation* is that it requires very limited screen space and reduces interaction as the text proceeds automatically. The two most well known formats for dynamic text presentation are *Leading* and *Rapid Serial Visual Presentation* (RSVP) [7]. Leading, or the Times Square Format, scrolls the text on one line horizontally across the screen whereas RSVP presents the text as chunks of words in rapid succession at a single visual location. RSVP seems to be more effective to use and it has proved to be just as fast as reading from a book or on a screen in several evaluations [3, 8, 10, 14]. RSVP originated as a tool for studying reading behavior, but has lately received more attention as a presentation technique with a promise of optimizing reading efficiency. The format is commonly claimed to reduce [3, 13, 14], or even eliminate [12, 18], the need for eye movements, which is assumed to increase reading speed and decrease task load. *Bailando* is an example of a program that makes use of dynamic text presentation in the RSVP format (Fig. 1 right) [10]. The width of the text presentation area is 25 characters with the text presented left justified in a 10-pt. sans-serif typeface; no legibility enhancing techniques are used. In order to support memory of spatial location while reading there is a progress bar, the inclusion of one has previously been found to increase the user preference for the RSVP format [14]. Bailando offers the user full control of the text presentation and it is possible to start/stop, increase/decrease speed, or go forward/backward at any time. The interface is designed for a PDA, but it could be ported to devices with much smaller displays.

3.3 Benchmarking the Page and RSVP Formats

In a previous usability evaluation performed on a PDA with 16 subjects, traditional text presentation in the Page format with Microsoft Reader was benchmarked against dynamic text presentation in the RSVP format with Bailando. The experiment was primarily designed to evaluate the effect of adaptation on RSVP, e.g. the adjustment of the text presentation speed to the time required for cognitive processing, but it was also interesting since it was the first evaluation of RSVP and traditional text presentation where all conditions were performed on a mobile device [10]. The results showed no significant differences in reading speed or comprehension between the Page and RSVP format. The results also showed that RSVP with adaptation could decrease NASA-TLX task load [4] ratings significantly for most factors, although it still remained higher for the RSVP formats compared to the Page format. Now, apart from the increase in task load, the RSVP format seems to be a viable alternative to traditional text presentation on mobile devices with a potential to increase readability on small screens. However, the question remains why RSVP increases task load, although it is assumed to decrease with reduced eye movements, and more importantly, how task load can be reduced. In order to learn more about this, and of how small screens affect readability in general, we decided to develop a tool that would enable us to monitor eye movements while reading on a mobile device.

4 Eye Movement Tracking

Using eye movements as a measure of readability connects well with our revised definition of readability, i.e. “the ease with which the reading process can proceed”. The main assumption being that eye movements conflicting with how we usually read can be seen as an indication of increased task load and decreased readability. The aim with the current study was to observe how the traditional and dynamic text presentation affects readability in terms of comprehension score, reading speed, task load rating, and eye movements. We wanted to compare the conditions that fared best in the previous evaluation, e.g. traditional text presentation in the Page format with Microsoft Reader and dynamic text presentation in the RSVP format with Bailando using content adaptation [10]. Moreover, we wanted to record eye movements and evaluate all conditions when reading on a mobile device.

We used the IOTA XY-1000 system for eye movement detection. The system uses infra-red (IR) light to detect eye movements and consists of a pair of goggles and a small processing unit (Fig. 2). In the goggles there are eight IR sensors, four for each eye, which detects horizontal and vertical eye movements. The processing unit is connected to a PC running the Orbit eye trace program, which converts the eye movements into horizontal and vertical coordinates and records them. The benefit of the system is that it is comfortable to wear, not invasive, and can record eye movements with a frequency of up to 1000 Hz. The downside is that the recordings are affected by head movement.

Fig. 2. The IOTA XY-1000 system with goggles (left) and processing unit (right)

Before recording any eye movements the system has to be calibrated. This is extremely important as the output data is a set of eye coordinates relative to a position that must be known. In order to be able to calibrate and align the eye movement recording system with the text presentation on the iPAQ, we developed a program called BAICOM. The program automatically sets up an eye movement recording session, maintains synchronization with the mobile device, and enables monitoring of the recording throughout the session. Equally important is to align the eye movement recording with the stimulus, in this case the text presentation, once the eye movement system is calibrated. These positions are then used as reference when analyzing the eye movement recordings. The eye movement recordings are analyzed using two programs. The first program, JR, is used to translate the recorded eye movement coordinates into real distances from the center of the screen measured in degrees. The second program, Eyealign, is used to classify and analyze the eye movements detected by the JR program on basis of their amplitude, duration, and co-occurrence.

5 The Eye Movement Study

The aim with the study was to see how the traditional Page format and the dynamic RSVP format affected the ability to read on a mobile device. It was important that the same device was used for all conditions since the look and feel of the hardware was likely to bias the assessment.

5.1 Method

A balanced within-subject repeated-measurement experimental design was employed. Two conditions were formed where each subject read one text using either presentation format. The conditions were balanced against presentation order and texts, thus generating four combinations, which each were repeated four times yielding sixteen experimental sessions. One subject was assigned to each of the sixteen sessions at random**.** The following null hypotheses were set for using both formats:

- No difference in Reading Speed
- No difference in Comprehension
- No difference in Task Load
- No difference in Eye Movements

The hypotheses were tested in the SPSS V11.5 software using the repeated-measurement General Linear Model (GLM). The significance level was set to 5% and the level of multiple comparisons was Bonferroni adjusted.

Subjects. Sixteen subjects (eight males and eight females; mean age: 28) participated in the experiment. All stated that they were computer literate and eight had some previous experience of using a PDA. Eleven of the subjects needed visual correction (Sphere >± 0.50 D). All subjects had good stereo vision (TNO), six had left eye dominance, nine had right eye dominance, and one showed no dominant eye.

Texts. Two Swedish fiction texts of similar length (~2500 words) and difficulty (LIX 30) were chosen to be included in the experiment. The text difficulty was measured with LIX [1], a readability rating developed for Swedish texts that is comparable to the Flesh index [19] for English. Two similarly difficult but shorter texts (~500 words) were also used for training.

Apparatus. All texts were presented on a PDA (Compaq iPAQ 3630). The MS Reader program was used for text presentation in the Page format and the Bailando prototype was used for the RSVP text presentation. While reading, the subjects wore a pair of infrared eye tracking goggles (IOTA XY-1000). The goggles were connected to a PC running the Orbit eye trace system recording horizontal and vertical eye movements. The eye movement system was controlled and monitored by the experimenter throughout the experiment. Before and after reading each text a calibra-

tion and synchronization was performed. Subjects that used glasses had replacement lenses of the same strength in the goggles. The subjects that used contact lenses kept them on during the recordings.

Setting. The experiment took place in a dedicated eye movement laboratory equipped with all necessary monitoring and recording facilities. While reading the subject was seated in a comfortable chair with the head held in a fixed position by an adjustable kin support. The viewing angle and distance to the screen of the iPAQ was kept constant for all measurements (40 cm). Although a fixed head is not natural while reading, realism must be sacrificed for the sake of reliable experimental data (Fig. 3).

Fig. 3. Setup of the experiment with the subject to the left and the experimenter to the right

Instructions. Before the experiment, each subject was given written instructions that pointed out that it was the different text presentation formats and not the individual performance that was being tested. The instructions emphasized that the subjects should try to read as they *normally* would in any everyday situation. Each subject was encouraged to select a comfortable reading speed. Each subject was also told that they could halt the experiment at any time if they felt uncomfortable.

Procedure. Each subject first participated in two training sessions. First, they read a training text using the MS Reader and then they read a similar text via RSVP. The eye tracking equipment was used under these conditions in order to adjust the system to each individual and make them used to it. Each subject then read one of the texts presented either in the traditional page format or via RSVP. After having read the first text, the subject answered a set of inventories. If the first text was read using the traditional page format, the second text was read via RSVP, and vice versa. The same set of inventories was administered after the second text.

Inventories. After each session, there were two inventories to fill in. The first was a comprehension test made up of ten multiple-choice questions with three alternatives. The second inventory was the NASA-TLX Task Load Index [4], which was administered to check Mental, Physical, and Temporal demands, as well as Performance, Effort, and Frustration levels. The NASA-TLX Task Load Index inventory was chosen as a measure of cognitive demand since the results would then be comparable to previous evaluations where the measure was used [3, 10].

5.2 Experimental Results

All subjects completed the experiment and there were few problems with understanding what to do or how to do it.

Reading Speed. Reading speed was calculated as words read per minute (wpm), based on the total time it took for the subjects to read a text (e.g. excluding the calibration but including all kind of interruptions like pauses, page turns, speed changes etc). The Page format increased reading speed some, but the null hypothesis regarding no difference in reading speed between the conditions was kept (Table 1).

Comprehension. Comprehension was computed as percent of correctly answered multiple-choice questions. For each text, there were ten questions with three choices. The Page format gave the best result, but the differences between the conditions were small. The null hypothesis regarding no difference in comprehension between the conditions was kept (Table 1).

Table 1. Reading Speed and Comprehension

Condition	*Reading Speed (wpm)*		*Comprehension (%)*	
	Avg.	Std dev.	Avg.	Std dev.
Page format	216.9	78.7	78.1	14.7
RSVP format	191.9	45.1	74.4	20.0

Task load. Task load was enumerated as percent of millimeters to the left of the tick mark on a 120-mm scale. The factors were not rated within each other. The null hypothesis regarding no difference in task load between the conditions was rejected as there was a significant difference in Temporal demand ($F[1,15] \geq 25.4$, $p \geq 0.001$). Pairwise comparisons revealed that the use of RSVP format resulted in significantly higher ($p \leq 0.001$) Temporal demand compared to using the Page format (Fig. 4).

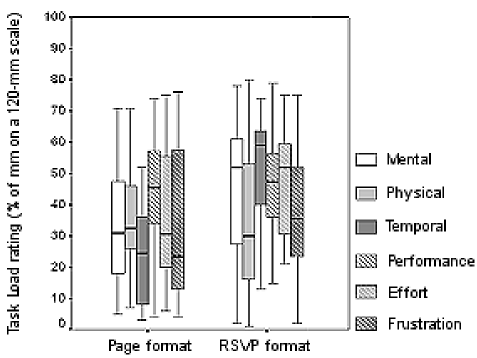


Fig. 4. NASA-TLX Task Load Index ratings. Lower ratings are better

5.3 Eye Movement Analysis

An eye movement was defined as a continual change in the recording with duration of more than 10 ms independently detected in each of the four channels (e.g. horizontal and vertical movements for both left and right eye). Eye movements were categorized according to their function when reading as: saccades, regressions, forward sweeps, return sweeps, stray sweeps, or eye blinks.

Saccades. A saccade was defined as a forward directed eye movement of 4 degrees or less with no simultaneous vertical movement independently detected for left and right eye. A threshold of 4 degrees corresponds to a visual span of approximately 20 characters. The null hypothesis regarding no difference was rejected as there was a significant difference ($F[1,15] \geq 43.7$, $p \leq 0.001$). Pairwise comparisons showed that the RSVP format significantly decreased the number of saccades, for both left ($p \leq 0.001$) and right ($p \leq 0.001$) eye, compared to the Page format (Table 2).

Regressions. A regression was defined as a backward directed eye movement of 4 degrees or less with no simultaneous vertical movement independently detected for left and right eye. The null hypothesis regarding no difference was rejected as there was a significant difference ($F[1,15] \geq 31.3$, $p \leq 0.001$). Pairwise comparisons showed that the RSVP format significantly increased the number of regressions, for both left ($p \leq 0.006$) and right ($p \leq 0.001$) eye, compared to the Page format (Table 2).

Forward sweeps. A forward sweep was defined as a forward directed eye movement exceeding 4 degrees without vertical movement, or a forward directed eye movement of less than 4 degrees with simultaneous vertical movement, both independently detected for left and right eye. The Page format was found to increase the number of forward sweeps, but not significantly. The null hypothesis regarding no difference was kept (Table 2).

Table 2. Saccades, Regressions, and Forward sweeps detected per minute

Condition	*Saccades*		*Regressions*		*Forward sweeps*	
	Left	Right	Left	Right	Left	Right
Page format	73.9	84.4	30.4	26.5	8.7	5.5
RSVP format	50.6	57.7	39.1	39.7	4.0	2.0

Return sweeps. A return sweep was defined as a backward directed eye movement exceeding 4 degrees without vertical movement, or a backward directed eye movement of less than 4 degrees with simultaneous vertical movement, both independently detected for left and right eye. The null hypothesis regarding no difference was rejected as there was a significant difference ($F[1,15] \geq 24.5$, $p \leq 0.001$). Pairwise comparisons showed that the Page format significantly increased the number of return sweeps, for both left ($p \leq 0.001$) and right ($p \leq 0.001$) eye (Table 3).

Stray sweeps. A stray sweep was defined as a horizontal eye movement of 4 degrees or more with simultaneous vertical movement in any direction, independently detected for left and right eye. The null hypothesis was rejected, as there was a significant difference ($F[1,15] \geq 12.6$, $p \leq 0.003$). Pairwise comparisons showed that the Page format significantly increased the number of stray sweeps, for both left ($p \leq 0.003$) and right ($p \leq 0.015$) eye, compared to the RSVP format (Table 3).

Eye blinks. An eye blink was defined as one or more vertical and horizontal eye movements occurring simultaneously in both eyes within a timeframe of one second (e.g. covering the closure and opening of the eyelids). The null hypothesis regarding no difference between the conditions was kept as there were no significant differences (Table 3).

Table 3. Return sweeps, Stray sweeps, and Eye blinks detected per minute

Condition	*Return sweeps*		*Stray sweeps*		*Eye blinks*	
	Left	Right	Left	Right	Left	Right
Page format	19.5	23.5	3.2	3.1	9.6	9.9
RSVP format	6.5	5.6	0.7	0.3	9.1	9.5

6 Discussion

Initially we stated that we needed to change the way we present information on small screens in order to improve readability on mobile devices. Our findings show that the traditional Page format actually offers better readability than the dynamic RSVP format on the PDA used in the experiment. Do we now really need to change the way we present information on mobile devices to achieve better readability? We believe the answer to be yes. The reason for this is that the RSVP format is likely to offer just as good readability on devices with far smaller screens than the PDA used in the experiment, whereas readability of the Page format is likely to decrease with diminishing screen size. The results obtained for Reading speed, Comprehension, and Task load were similar to those from the previous evaluation [10]. The largest differences between the conditions were found in the eye movement analysis and we will therefore limit the discussion to these results and their implications for further work.

Using the RSVP format decreased the overall number of eye movements with around one third, but the results show that there is still a considerable amount of eye movement activity going on. As mentioned earlier, it is commonly claimed that the RSVP format reduces [3, 13, 14], or even eliminate [12, 18], the need for eye movements. Our findings show that using RSVP with a 25 character text display indeed reduces eye movements, but is far from eliminating them. The eye movements resulting from using the Page format closely resembles the eye movements performed when reading on paper. The saccade/regression ratio for the Page format was around 3 to 1, whereas the ratio was closer to 1 to 1 for the RSVP format. These figures

should be compared to a ratio of around 5 to 1 for reading on paper [6]. A regression is typically executed when something has gone amiss in the reading process, and in this sense a large amount of regressions can be an indication of lower readability. Given that the Page format offers a much larger text presentation area, the higher amount of forward sweeps, return sweeps, and stray sweeps compared to the RSVP format is not too surprising. All page return sweeps, e.g. going to the top of the next page, were also classified as stray sweeps. That the RSVP format yields forward and return sweeps may seem unexpected as the text is presented only on one line, but most of these are eye movements exceeding four degrees without vertical movement (e.g. long saccades or regressions). We expected eye blinks to increase with cognitive demand, but this was not the case.

The advantage of the RSVP format was originally presumed to be the elimination of eye movements, which would lead to a possible reduction in cognitive load [12]. Our results show that the RSVP format does not eliminate eye movements, although it does reduce them. The reduction does however not seem to reduce cognitive load. In fact, it rather seems to increase cognitive load. The reason for this may be the increase in regressions, which can be seen as an indication of when the reading process has not proceeded with ease. These empirical findings contradict the theoretical basis of RSVP, which means that we may have to reconsider the format. A dynamic text presentation format like RSVP should maybe not try to reduce eye movements, but rather try to stimulate an eye movement pattern similar to when reading on a paper page. What if small changes to the position of the RSVP text chunks could reduce task load further? A dynamic text presentation format that predicts when the reader is about to move the eyes and where to display the next text segment intuitively seems to relate closer to how the reading process works. Using eye movement tracking as a tool when evaluating how well different text presentation formats support reading seems to be a valuable approach to learn more about this, and of how to improve readability on mobile devices in the future.

7 Conclusions

We have used eye movement tracking as a tool to learn more about how different methods of text presentation affect reading on a mobile device. The results from the eye movement study, where the traditional Page format was compared to the dynamic RSVP format, demonstrate the value of the approach. In terms of reading speed and comprehension, we found no significant differences in reading efficiency, although task load was higher for the RSVP format. RSVP was found to decrease eye movements compared to the Page format, but was far from eliminating them. In fact, although the RSVP format decreased the amount of saccades it was found to increase the amount of regressions. The increase in regressions may explain the higher task load as it can be seen as an indication of lower readability. Our empirical findings disprove the theoretical assumption of the RSVP format, that suppressing eye movement reduces cognitive load. Instead we propose that a dynamic text presentation should try to stimulate eye movements similar to how we are accustomed to read.

References

1. Björnsson, C.H. (1968). *Läsbarhet*. Stockholm: Liber.
2. Ericsson, T., Chincholle, D. and Goldstein, M. (2001). Both the device and the service influence WAP usability. *Proceedings of IHM-HCI2001, Usability in Practice,* J. Vanderdonckt, A. Blandford and A. Derycke (Eds.), 10-14 September, Lille, France, 79-85.
3. Goldstein, M., Sicheritz, K. and Anneroth, M. (2001). Reading from a small display using the RSVP *technique. Proceedings of Nordic Radio Symposium*, *NRS01*. Nynäshamn, Sweden.
4. Hart, S.G. and Staveland, L.E. (1988). Development of Nasa-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*, P.A. Hancock and N. Meshkati (Eds.). Elsevier: North-Holland.
5. Hill, B. (2001). *The Magic of Reading*. Available at: http://www.microsoft.com/reader/includes/TheMagicofReading.lit (November 2003).
6. Just, M. A., and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329-354.
7. Mills, C.B. and Weldon, L.J. (1987). Reading text from computer screens. *ACM Computing Surveys*, 19(4), 329-358.
8. Muter, P. (1996). Interface design and optimization of reading of continuous text. *Cognitive Aspects of Electronic Text Processing*, H. van Oostendorp and S. de Mul (Eds.). Norwood, NJ: Ablex, 161-180.
9. Muter, P. and Maurutto, P. (1991) Reading and skimming from computer screens and books: The paperless office revisited? *Behaviour & Information Technology*, 10(4), 257-266.
10. Öquist, G. and Goldstein, M. (2003). Towards an improved readability on mobile devices: evaluating adaptive rapid serial visual presentation. *Interacting with Computers*, 15(4), 539-558.
11. Paulson, L.J. and Goodman, K.S. (2000). *Influential studies in eye-movement research*. International Reading Association. Available at: http://www.readingonline.org (November 2003).
12. Potter, M. C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. *New Methods in Reading Comprehension Research*, D. E. Kieras and M. A. Just (Eds.). Hillsdale, N.J: Erlbaum.
13. Proctor, R. W. and Proctor, J.D. (1997). Sensation and Perception. *Handbook of Human Factors and Ergonomics*, G. Salvendy (Ed.), Second Edition. New York, NY: Wiley.
14. Rahman, T. and Muter, P. (1999). Designing an interface to optimize reading with small display windows. *Human Factors*, 1(1), 106-117.
15. Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422.
16. Reichle, E.D., Rayner, K. and Pollatsek, A. (2000). *Comparing the E-Z Reader Model to Other Models of Eye Movement Control in Reading*. Cog Prints Archives. Available at: http://cogprints.soton.ac.uk/ (November 2003).
17. Robeck, M.C. and Wallace, R.R. (1990). *The Psychology of Reading: An Interdisciplinary Approach*. Hillsdale, NJ: Erlbaum.
18. Rubin, G. S., and Turano, K. (1992). Reading without saccadic eye movements. *Vision Research*, 32, 895-902.
19. Tekfi, C. (1987). Readability formulas: An Overview. *Journal of Documentation*, 43(3), 261-73.
20. Weiss, S. (2002). *Handheld Usability*. New York, NY: Wiley and Sons.

Tilt-Based Automatic Zooming and Scaling in Mobile Devices – A State-Space Implementation

Parisa Eslambolchilar[1] and Roderick Murray-Smith[1,2]

[1] Hamilton Institute, National University of Ireland, NUI, Maynooth, Co.Kildare, Ireland
parisa.eslambolchilar@may.ie

[2] Department of Computing Science, Glasgow University, Glasgow G12 8QQ, Scotland
rod@dcs.gla.ac.uk

Abstract. We provide a dynamic systems interpretation of the coupling of internal states involved in speed-dependent automatic zooming, and test our implementation on a text browser on a Pocket PC instrumented with an accelerometer. The dynamic systems approach to the design of such continuous interaction interfaces allows the incorporation of analytical tools and constructive techniques from manual and automatic control theory. We illustrate experimental results of the use of the proposed coupled navigation and zooming interface with classical scroll and zoom alternatives.

1 Introduction

Navigation techniques such as scrolling (or panning) and zooming are essential components of mobile device applications such as map browsing and reading text documents, allowing the user access to a larger information space than can be viewed on the small screen. Scrolling allows the user to move to different locations, while zooming allows the user to view a target at different scales. However, the restrictions in screen space on mobile devices make it difficult to browse a large document efficiently. Using the traditional scroll bar, the user must move back and forth between the document and the scroll bar, which can increase the effort required to use the interface. In addition, in a long document, a small movement of the handle can cause a sudden jump to a distant location, resulting in disorientation and frustration.

Speed-dependent automatic zooming is a relatively new navigation technique [7, 8, 14, 22, 25, 26] that unifies rate-based scrolling and zooming to overcome these limitations. The user controls the scrolling speed only, and the system automatically adjusts the zoom level so that the speed of visual flow across the screen remains constant. Using this technique, the user can smoothly locate a distant target in a large document without having to manually interweave zooming and scrolling, and without becoming disoriented by extreme visual flow.

In this paper we demonstrate that, as suggested by Igarashi and Hinckley [14], SDAZ is well suited to implementation on mobile devices instrumented with tilt sensors, which can then be comfortably controlled in a single-handed fashion. We also describe an alternative stylus controlled implementation for the PocketPC. A further contribution is the use of a state-space formulation of speed dependent zooming,

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 120–131, 2004.

which we believe is a promising reformulation of the technique, which opens the path to the use of analytic tools from optimal and manual control theory.

2 Speed-Dependent Automatic Zooming – A Brief Review

Several techniques have been proposed to improve the manipulation of scroll bars [14, 19]. They allow the user to control scrolling speed, enabling fine positioning in large documents. LensBar [18] combines these techniques with interactive filtering and semantic zooming, and also provides explicit control of zooming via horizontal motion of the mouse cursor. A rate-based scrolling interface is described in [29] that maps displacement of the input device to the velocity of scrolling.

Zoomable user interfaces, such as Pad and Pad++ [4], use continuous zooming as a central navigation tool. The objects are spatially organized in an infinite two-dimensional information space, and the user accesses a target object using panning and zooming operations. A notable problem with the original zoomable interfaces is that they require explicit control of both panning and zooming, and it is sometimes difficult for the user to coordinate them. The user can get lost in the infinite information space [16]. Bimanual approaches also exist, such as that of Guiard *et al.* [11] where a joystick in one hand controlled zoom level, and a mouse in the other provided navigation. They showed that by using zooming interfaces, bit rates far beyond those possible in physical selection tasks become possible.

Information visualization techniques, such as Fisheye Views [9, 12], Perspective Wall [17], and the Document Lens [21] also address the problem of information overload by distorting the view of documents. The focused area is magnified, while the non-focused areas are squashed but remain in spatial context. The user specifies the next focal point by clicking or panning. Van Wijk derived an optimal trajectory for panning and zooming in [24], for known start and end points.

The particular input device used can also influence the effectiveness of rate control. An experiment on 6 DOF input control [29] showed that rate control is more effective with isometric or elastic devices, because of their self-centring nature. It is also reported that an isometric rate-control joystick [2] can surpass a traditional scroll bar and a mouse with a finger wheel [29]. Another possibility is to change the rate of scrolling or panning in response to tilt, as demonstrated by Rekimoto [20] as well as Harrison *et al.* [13], suitable for small screen devices like mobiles phones and PDAs.

A common problem with scrolling and zooming interfaces is that when users are zoomed out for orientation, there is not enough detail to do any ‘real work’. When they are zoomed in sufficiently to see detail, the context is lost. To reduce this problem, multiple windows can be provided, each with pan and zoom capability. Although this is reasonable for small information spaces, the many windows required by large spaces often lead to usability problems due to excessive screen clutter and window overlap. An alternative strategy is to have one window containing a small overview, while a second window shows a large more detailed view [3, 10]. The small overview contains a rectangle that can be moved and resized, and its contents are shown at a larger scale in the large view. This strategy, however, requires extra space for the overview and forces the viewer to mentally integrate the detail and context views. An operational overhead is also required, because the user must regularly move the mouse between the detail and context windows.

Speed-dependent automatic zooming (SDAZ) is a navigation technique first proposed by Igarashi & Hinckley [14]. It couples rate-based scrolling with automatic zooming to overcome the limitations of typical scrolling interfaces and to prevent extreme visual flow. This means that as a user scrolls faster the system automatically zooms out, providing a constant information flow across the screen. This allows users to efficiently scroll a document without having to manually switch between zooming and scrolling or becoming disoriented by fast visual flow, and results in a smooth curve in the space-scale diagram. In traditional manual zooming interfaces, the user has to interleave zooming and scrolling (or panning); thus the resulting pan-zoom trajectory forms a zigzag line. Cockburn *et al.* [7, 8, 22, 25, 26] presented further developments, with a usability study of performance-improved SDAZ prototypes.

3 Dynamics and Interaction

In this paper we use systems of differential equations to describe the interaction between user and computer. Skeptics might question this *"Why introduce dynamics, when dynamic systems tend to be more difficult to control than static ones? Vehicle control systems tend to go to great trouble to hide the underlying dynamics of the vehicle from the driver."*

We explicitly include dynamics because we can only control what we can perceive, and while, in principle, we can navigate instantly in an arbitrary information space, given a static interaction mechanism (e.g. clicking on a scroll bar), if we are dependent on feedback to be displayed while pursuing our goals, there will be upper limits on the speed at which the display can change. This is especially true in cases where there is uncertainty in the user's mind about where to go, and when they have the option to change their goal on route, as more information becomes available. In order to cope with this, interface designers have a long history of hand-crafting transition effects in a case-by-case manner. Nonlinear mouse transfer functions are long-established examples of finely-tuned dynamic systems driven by user input.

One of our long-term goals is to investigate whether describing the dynamics of interaction using the tools of control engineers allows us a more consistent approach to analyzing, developing and comparing the *'look-and-feel'* of an interface, or in control terms, the *'handling qualities'*. Control synthesis often focuses on analysis of coupling among system states. Speed-dependent zooming is an obvious example of this, but if we generalize the approach to other interaction scenarios, with possibly a larger number of interacting states/inputs, we will require more general methods to analyse the consequences of coupling effects. Control methods are likely to be especially important for design for mobile devices, where sensor noise, disturbance rejection, sensor fusion, adaptive self-calibration and incorporating models of human control behaviour are all important research challenges.

In cases such as the use of accelerometers as input devices, the direct mapping of acceleration in the real world to acceleration in the interface provides an intuitive mapping, which also suggests a range of other affordances, especially for multi-modal feedback, which can then be utilized by interface designers. Real-world effects such as haptic feedback of springs, or friction linked to speed of motion are easy to reproduce in a dynamic system, and we can choose to explicitly use these features to design the system to encourage interaction to fall into a comfortable, natural rhythm.

Furthermore, the act of performing a continuous input trajectory to achieve a goal, creates proprioceptive feedback for the user which can then be associated with that particular task. The mechanisms of gesture recognition can be ‘opened up’ and explicitly made visible *during* the motion, to provide a link for the user between the control input and the task completion. We describe a probabilistic, audio/vibrotactile approach to this in [28], which can ease learning and reduce frustration.

The use of dynamic models of interaction allows *intelligent interaction*, if the handling qualities of the dynamics of the interface are adapted depending on current *inferred* user goals. Using this approach, actions require less effort, the more likely the system’s interpretations of user intentions, equivalent to a fewer bits from the user, in communication terms. This was used by Barrett *et al.* in [2], and we used this approach for text entry in Williamson & Murray-Smith [27], and the approach can be linked to methods which adapt the control-to-display ratio, such as Blanch *et al.* [5] in classical windows interfaces. These approaches, which work with relative input mechanisms, cannot be used if we use static mappings, such as a stylus touching an explicit point on the screen.

4 Speed-Dependent Automatic Zooming on a Mobile Device

Implementing the SDAZ technique on a mobile device with inertial sensing allows us to investigate a number of issues: the use of single-handed tilt-controlled navigation, which does not involve obscuring the small display; the usability consequences of tilting the display; the relative strength of stylus-based speed-dependent zooming, compared to mouse and tilt-based control, and combinations of stylus, and tilt-based control. If successful, the user should be able to target a position quickly without becoming annoyed or disoriented by extreme visual flow, and we want the technique to provide smooth transitions between the magnified local view and the global overview, without the user having to manually change the document magnification factor.

Fig. 1a **Fig. 1b**

Fig. 1. PocketPC and accelerometer attached to serial port (1a). Screen shots of the document browser (1b). The left picture shows a red box moving rapidly over the picture, the middle picture shows the user has found the picture and landing there, and right picture shows the zoomed-in picture.

4.1 Hardware/Software Environment

We implemented this method using Embedded Visual C++ on an HP 5450 Pocket PC (Figure 1). Here, tilting the device moves the zooming-window. The accelerometer (Xsens P^3C, 3 degree-of-freedom linear accelerometer) attached to the serial port of the Pocket PC provides the roll and pitch angles.

4.2 Design and Implementation of Speed-Dependent Automatic Zooming

State space modelling is a well-established way of presenting differential equations describing a dynamic system as a set of first-order differential equations. There is a wealth of knowledge and analysis techniques from systems theory, including designing estimators and controllers for multi–input–multi–output systems, optimal control, disturbance rejection, stability analysis and manual control theory [6]. State-space modelling allows us to model the internal dynamics of the system, as well as the overall input/output relationship as in transfer functions, so this method is an obvious candidate for the representation of the coupling between the user's speed with zoom level. There are many advantages to modelling systems in state space, especially for multivariable problems, where the matrix formulation is particularly useful for analysis purposes.

4.2.1 State Space Model

For an introduction to the basic ideas, see any introductory control theory book, e.g. [1,6]. The generic form for the state equations is given by equation (1)

$$\begin{aligned} \dot{X} &= f(x) + g(u) \\ \dot{Y} &= h(x) \end{aligned} \tag{1}$$

where f(x), g(u) and h(x) can be nonlinear functions, and where $X(t)$ is an $n \times 1$ state vector where n is the number of states or system order, $U(t)$ is a $r \times 1$ input vector where r is the number of input functions, and $Y(t)$ is a $p \times 1$ output vector where p is the number of outputs. The more specific case of a linear system, (2)

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{2}$$

where A is an $n \times n$ square matrix called the *system matrix,* B is an $n \times r$ matrix called the *input matrix,* C is a $p \times n$ matrix called the *output matrix and* D is a $p \times r$ matrix which represents any direct connection between the input and output.

4.2.2 Coupling the User's Velocity with the Zoom-Level

In this section we show how an SDAZ-like approach couples the user's motion with the zoom-level. The inputs to the system are the tilting angles measured using an accelerometer attached to the serial port of PDA, and in a second experiment the stylus position on the PDA touch screen. The state variables chosen are $x_1(t)$ for position, $x_2(t)$ for speed of scroll and $x_3(t)$ for zoom, and the state equations are:

$$x_2(t) = V = \Delta x_1 \tag{3}$$

$$x_3(t) = Z = f(x_1, x_2, u) \tag{4}$$

So the zoom-level is a function of position, velocity and tilting angle. An initial suggestion is to reproduce the standard second-order dynamics of a mass-spring-damper system, in the hope that giving the scrolling movement and zoom level some inertia will provide a physically intuitive interface. The first time-derivative of the state equations can be written as below, as a linearization of the system at a given velocity and zoom:

$$\dot{x}_1(t) = V = x_2(t) \tag{5}$$

$$\dot{x}_2(t) = \dot{V} = \frac{-R}{M} x_2(t) + \frac{k}{M} u(t) \tag{6}$$

$$\dot{x}_3(t) = \dot{Z} = \frac{-b}{M} x_2(t) + \frac{-R}{M} x_3(t) + \frac{a}{M} u(t) \tag{7}$$

The standard matrix format of these equations is:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & \frac{-R}{M} & 0 \\ 0 & \frac{-b}{M} & \frac{-R}{M} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{k}{M} \\ \frac{a}{M} \end{bmatrix} u(\mathbf{t}) \tag{8}$$

This shows how a single-degree of freedom input can control both velocity and zoom-level. The non-zero off-diagonal elements of the *A* matrix indicate coupling among states, and the *B* matrix indicates how the inputs affect each state. This example could be represented as having zoom as an output equation, rather than state, and the coupling between zoom and speed comes only through the *B* matrix, which is not particularly satisfying. However, this paper is intended as an initial exploration of the area, and as more interesting behaviour can be obtained by fully interacting nonlinear equations, such as those elegantly derived by van Wijk in [24], we have left it in this format. In the experiments, *R=1, M=1, k=1 and b=0*, but we also experimented with varying the parameters, essentially including nonlinearities by a function relating velocity with zoom factor, as will be discussed in the next section. We include saturation terms for maximum and minimum zoom levels, and there can be specific rules for behaviour at the limits associated with the start and end of the document. For nonlinear functions we can locally linearise around any given state [*x v z*] leading to time-varying matrices *A(t),B(t).* We can analytically investigate the local dynamics for different operating points by, for example, looking at the eigenvalues of the *A* & *B* matrices to check for oscillatory (eigenvalues are complex conjugate pairs) or unstable behaviour (real part of eigenvalue in right half plane – i.e. positive). For more background see any control textbook (e.g. [1, 6]). Importantly, the system itself might be stable, but when coupled with the time delay and lead-lag-dynamics of typical human control behaviour, the combined closed loop system might be unstable, as in pilot-induced oscillations in aircraft control [15,23].

The dynamic systems implementation allows us to deviate from a static link between speed and zoom level. In this paper, our basic assumption is that zoom should lead speed when speed increases, in order to avoid extreme visual flow. Zoom should, however, lag speed when $|v|$ decreases, to allow the user to slow down but still maintain the overview. This also allows, for example, the user to zoom out, without changing position in the document, by repeated positive and negative acceleration.

In order to move more rapidly through the document at high levels of zoom, in this paper, we adapted B by making 'a' in eqn. (8) a function of velocity. When speed is above the dead-zone threshold (here set to 0.1), $a = 3$ but below this threshold $a=0$. We wish to avoid rapid drop effects when user changes direction. To achieve this, we set $a=a*0.2$, when the sign of velocity and input differ. For practical implementation on a PDA we converted the continuous-time system to a discrete-time one [1], with sampling time h, which involves the evaluation of a matrix exponential, $\Phi = e^{Ah}, \quad \Gamma = \int_0^h e^{As} ds B$.

$$x(kh+h) = \Phi x(kh) + \Gamma u(kh) \tag{9}$$
$$y(kh) = Cx(kh) + Du(kh)$$

A phase plane figure shows an example of a trajectory through this state-space for the SDAZ implementation on the Pocket PC (Figure 2). This gives some insight into the transient dynamics of large and small translations of position through the document.

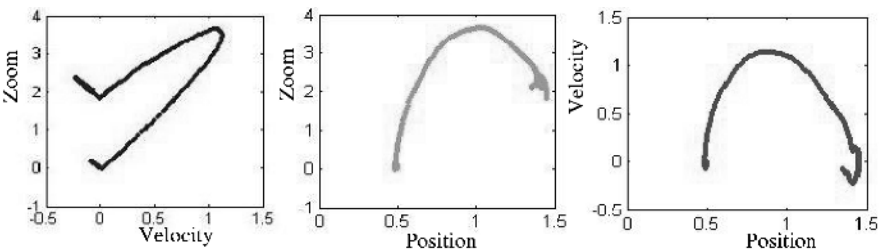


Fig. 2. Phase plane trajectories showing velocity against zoom (left), zoom-level against position (centre) and velocity against position (right), from a record of participant browsing a long document on the PocketPC.

4.2.3 Control Mode

We can now introduce transitions among control modes which alter the dynamics and the way user inputs are interpreted. A simple example of this approach uses state feedback to augment control behaviour, by making the state move towards some reference value r, we can create a control law $u = L(r - x)$, such that the new state equations are

$$\begin{aligned} \dot{x} &= Ax + Bu = Ax - BLx + BLr \\ &= (A - BL)x + BLr \end{aligned} \tag{10}$$

such that the system dynamics have changed from A to $(A\text{-}BL)$. In the SDAZ implementation in this paper, we switched from tilt-angle as acceleration, to tilt angle to indicate desired velocity, as soon as the speed passed the threshold at which zooming started. This made it easier for users to find and maintain a comfortable zoom level. Other similar examples can be created, where the interpretation of sensor inputs and their significance for control can adapt to context. Including position control, for example, would allow the user to tap on the screen to specify a goal, which is then dynamically acquired. While on route to that goal, the user changes their mind, they can break out and switch again to velocity control.

4.2.4 Calibrating SDAZ and the State Space Approach

SDAZ has many parameters that can be tuned, usually treated as a series of interacting, but essentially separate equations. The state-space formulation allows multiple variables, and derivative effects (e.g. position, velocity, acceleration) can be coupled with zoom level, without any further coding, by just changing the entries of the A matrix, simulating combinations of springs, masses and damping effects.

In SDAZ, the function linking zoom with velocity, $z = f(v)$, can be nonlinear, including threshold effects. Examples include linear, with thresholds, exponential, and 'modified exponential' [14,25]. Furthermore the document velocity $v=g(\delta)$ as a function of control input (mouse displacement, tilt-angle, or stylus displacement, depending on platform) tend to be static, linear, or piecewise linear functions [14, 25]. In the state-space representation, we need to reformulate these equations in terms of the time-derivatives of zoom and velocity, via the A and B matrices. For example, for ramp increases in speed, the modified exponential zoom-speed mapping corresponds to our suggestion of zoom leading speed, with the exponent being related to the difference between the time constants for zoom and speed.

To enhance the smoothness of the transition between the global overview and the magnified local view after a mouse button is pressed, Cockburn and Savage use a 'falling' speed, and Igarashi & Hinckley [14] place a limit on the maximum time-derivative of zoom, with similar effect. The falling rate was calculated using trial and error – if the rate was too fast, the user felt motion sickness and lost their place in the document, whereas it being too small led to a sluggish interface. This can be represented as a straightforward switch to a particular parameterization of the A matrix, which can be tuned to give an appropriate exponential decay in velocity or zoom.
Related problems include rapid zooming in and out when making a rapid change of direction [14]. In the state-space representation, dealing with these issues becomes a matter of tuning the dynamics of the system by changing the A matrix, to make, for example, the time-constants associated with the zoom level larger than that of the speed, for regimes where speed is dropping.

Gutwin [12], Igarashi & Hinckley [14] and Wallace [25] report the *hunting effect* problem when users overshoot the target due to the system zooming in as the user slows, the user then rapidly adjusts behaviour to compensate, which causes the system to zoom out again. One approach to this would be to switch to a 'diving' control mode if $dz/dt < z_{thresh}$, where $a=0$, preventing zooming increases, unless a major change in velocity, occurs, which would switch the control mode back to velocity control.

5 Example Application – Document Browser for a PDA

The document viewer was designed to use automatic zooming to browse PDF, PS and DOC files which had been converted to a image (PNG) file. BMP or PNG (Portable Network Graphics) files are more efficient, and have low rendering time. This increases the speed and smoothness of the browser, the implementation of which was simple but very efficient and smooth (although text tended to flicker during zooming because it was treated as a flat image). Equations (15) to (18) (previous section) show the formula used to calculate the relationship between the user's hand motion (tilting PDA) and the zoom level from the document.

For comparison we show trajectories of users using traditional scroll bars on the Pocket PC and a touch-screen based SDAZ implementation (Figure 3) for browsing a long document on PDA (Figure. 1b). The touch-screen based SDAZ and tilt-controlled SDAZ both use the same state-space model. The results in Figure 3 highlight the different navigation styles of the different interfaces, with the scroll bar approach using a number of rapid translations through the document to find a paragraph in bottom of the document, and no use of zooming for an overview, while the two SDAZ implementations had smoother navigation, which also included smooth changes in zoom level.

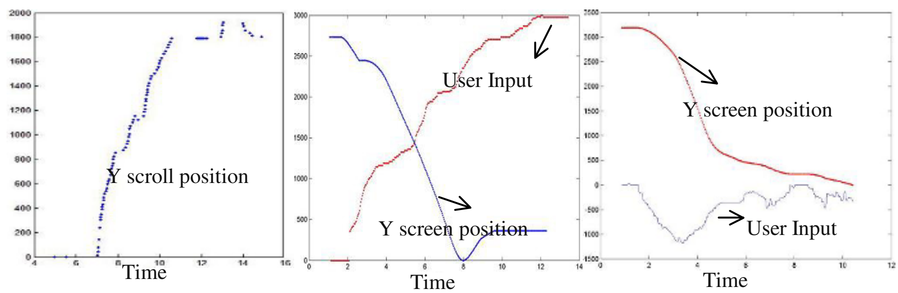


Fig. 3. Left picture shows the trajectory of one participant in using traditional scroll bars in browsing the long document, so *y* displacement is as long as the document. Middle picture shows the trajectory of the same participant in touch screen based SDAZ in browsing the long.

Users found the touch screen-based mechanism intuitive and easy to use for browsing. Figure 4 presents the system's inputs in three SDAZ applications to find the same paragraph used in scroll bar browser for tilt-based and touch screen controlled SDAZ. Also this figure presents an example run with tilt-based SDAZ, with augmented velocity control, as described in section 4.2.3, to browse the document to find 7 main headings. For comparison, the central plots in Figure 4 show tilt-based SDAZ without augmented velocity control on the same task, where fluctuations indicate that controlling the zoom level was difficult, and hunting behaviour appears when users tried to land on the targets (e.g. t=20,40, 85, in middle figures).

6 User Feedback

We asked five users from our research lab to work with the document browser using tilt-based SDAZ and touch screen-controlled SDAZ with and without augmented velocity control. Users who did the experiment without augmented velocity control suggested that adding a control option or a switch to control the zoom-level with velocity and tilting angles will make the system more comfortable to use. Most of them proposed if they could control level of zoom by tapping on the screen or pressing a key on PDA, the application would be easier to use.

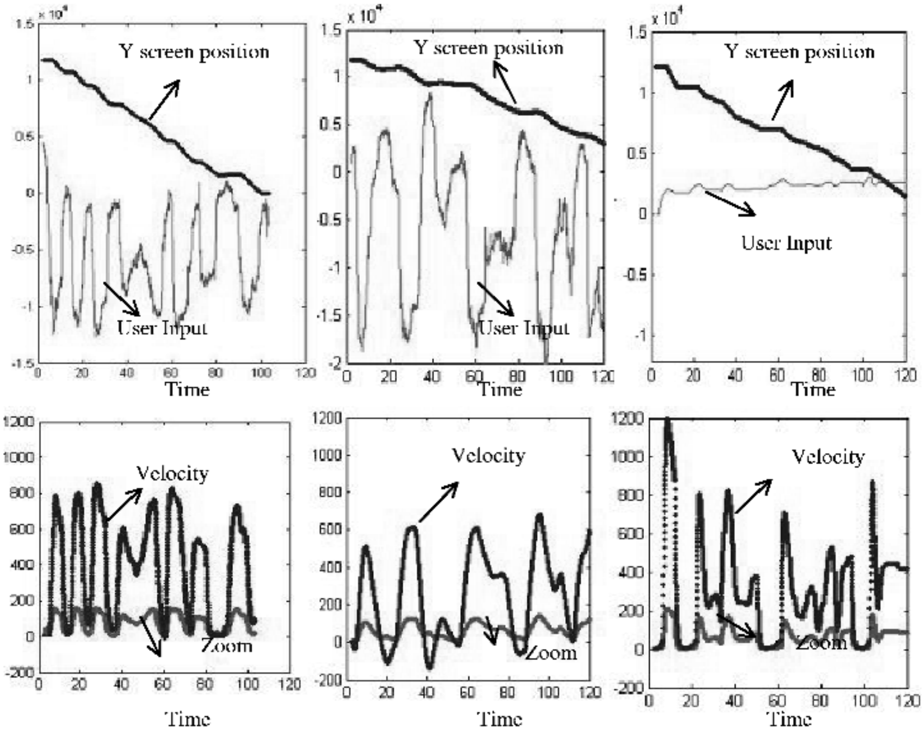


Fig. 4. Left picture tilt-based SDAZ with augmented velocity control, middle picture tilt-based SDAZ without augmented control and right picture touch-screen controlled SDAZ.

In contrast, users who did their experiments with augmented velocity control were satisfied with the application in both tilt-based and touch screen-controlled modes. Some users complained that with tilt input, they had to tilt the device to angles which caused irritating reflections from the PocketPC screen. Users in both groups, with and without augmented control, commented that if they were involved with other tasks, (like answering the phone, working with PC, etc.) they would prefer the touch screen-controlled SDAZ because they imagined it would be difficult to stay in the desired position in the document, with a tilt-based SDAZ. Although this was beyond the scope of our initial experiments, a key factor in the usefulness of tilt-based SDAZ will be the ease with which the user can toggle tilt-control on and off, during tasks.

7 Conclusions

We have presented a state-space, dynamic systems representation of the dynamic coupling involved in speed-dependent automatic zooming. We demonstrated the applicability of the approach by implementing a speed-dependent zooming interface for a text browsing system on a PDA instrumented with an accelerometer, and with stylus control. We illustrated the behaviour of the different interfaces by plotting their trajectories in phase space and as time-series.

Initial informal user evaluation of the implementation of SDAZ on a Pocket PC is positive, and users felt that this provided an intuitive solution to the problem of large documents and small displays. The tilt-controlled version can be used in a single-handed manner, without obscuring the screen, but because in the implementation tested, there was no toggle for tilt-control, users felt more comfortable with the stylus-controlled version.

This approach has the potential to provide a very general framework for development, analysis and optimisation of interfaces which induce complex, but convenient coupling among multiple states, in order to cope with few degrees of freedom in input. It opens up the dynamics of the 'look and feel' of mobile applications based on continuous control metaphors, to analysis and design techniques from automatic and manual control theory [15, 23].

Acknowledgements. The authors gratefully acknowledge the support of SFI BRG project *Continuous Gestural Interaction with Mobile devices*, Science Foundation Ireland grant 00/PI.1/C067, the *MAC network* - EC TMR grant HPRN-CT-1999-00107, and EPSRC grant GR/R98105/01.

References

[1] K. Åström, J., Wittenmark, B., *Computer controlled systems, Theory and Design*, 3rd ed. NJ: Prentice Hall, 1997.
[2] R. C. Barrett, E. J. Sleker, J. D. Rutledge, and R. S. Olyha, "The Negative Inertia: A dynamic pointing function," in CHI'95, conference companion, 1995, pp. 316-317.
[3] D. Beard and J. Walker, "Navigational techniques to improve the display of large two-dimensional spaces," *Behav. Inf. Tech*, vol. 9(6), pp. 451-466, 1990.
[4] B. Bederson and J. Meyer, "Implementing a Zooming User Interface: Experience Building Pad++," *Journal of Software - Practice and Experience*, vol. 28(10), pp. 1101-1135, 1998.
[5] R. Blanch, Y. Guiard, M. Beaudouin-Lafon, "Semantic Pointing: Improving target acquisition with Control-Display ratio adaptation," *Proc. of ACM Conference on Human Factors in Computing Systems*, pp. 519- 526, 2003.
[6] W. L. Brogan, *Modern Control Theory*, 3rd ed: NJ: Prentice Hall, 1991.
[7] A. Cockburn, J. Looser, and J. Savage, "Around the World in Seconds with Speed-Dependent Automatic Zooming," in Demonstration in the Proceedings of the ACM User Interface Software and Technology (UIST Conference Supplement), Vancouver, Canada, 2003, pp. 35-36.
[8] A. Cockburn and J. Savage, "Comparing Speed-Dependent Automatic Zooming with Traditional Scroll, Pan, and Zoom Methods," in People and Computers XVII: British Computer Society Conference on Human Computer Interaction, Bath, England, 2003, pp. 87-102.

[9] G. W. Furnas, "Generalized Fisheye Views," in Proceedings of CHI'86, 1986, pp. 16-23.

[10] G. W. a. B. B. B. Furnas, "Space-Scale Diagrams: Understanding Multiscale Interfaces," *Proceedings of CHI'95*, 1995.

[11] Y. Guiard, F. Bourgeois, D. Mottet, and M. Beaudouin-Lafon, "Beyond the 10-Bit Barrier: Fitts' Law in Multiscale Electronic Worlds," in Proceedings of IHM-HCI, 2001, pp. 573-587.

[12] C. Gutwin, "Improving focus targeting in interactive fisheye views," *Proc. of ACM Conference on Human Factors in Computing Systems (CHI'02), Minneapolis*, pp. p267-274, 2002.

[13] B. Harrison, K. P. Fishkin, et al., "Squeeze Me, Hold Me, Tilt Me! An Exploration of Manipulative User Interfaces," *Proceedings of CHI'98.*, 1998.

[14] T. Igarashi and K. Hinckely, "Automatic speed-dependent zooming for browsing large documents," in 13th Annual Symposium on User Interface Software and Technology, ACM UIST, San Diego, CA, 2000, pp. 139-148.

[15] R. J. Jagacinski and J. M. Flach, *Control Theory for Humans*: Lawrence Erlbaum Associates, 2003.

[16] S. Jul and G. Furnas, "Critical Zones in Desert Fog: Aids to Multiscale Navigation," in Proceedings of UIST '98, 1998, pp. 97-106.

[17] J. D. Mackinlay, G. G. Robertson, and C. K. Card, "The Perspective Wall: Detail and Context Smoothly Integrated," in Proceedings of CHI'91, 1991, pp. 173-179.

[18] T. Masui, "LensBar - Visualization for Browsing and Filtering Large Lists of Data.," in Proceedings of InfoVis'98, 1998, pp. 113-120.

[19] T. Masui, K. Kashiwagi, and G. R. Borden, "Elastic graphical interfaces for precise data manipulation," in CHI'95, Conference Companion, 1995, pp. 143-144.

[20] J. Rekimoto, "Tilting Operations for Small Screen Interfaces," in Proceedings of UIST, 1996, pp. 167-168.

[21] G. G. Robertson and J. D. Mackinlay, "The Document Lens," in Proceedings of UIST'93, 1993, pp. 101-108.

[22] J. Savage, "Speed-dependent automatic zooming," University of Canterbury, 2002. http://www.cosc.canterbury.ac.nz/research/reports/HonsReps/2002/hons_0208.pdf.

[23] T. B. Sheridan and W. R. Ferrell, *Man-Machine Systems: Information, Control, and Decision Models of Human Performance*: MIT Press, 1974.

[24] J. J. Van Wijk, "Smooth and efficient zooming and panning," *T. Munzner, S. North (eds.), Proceedings IEEE Symposium on Information Visualization (InfoVis'2003)*, pp. 15-22, 2003.

[25] A. Wallace, "The Calibration and Optimisation of Speed-Dependent Automatic Zooming," University of Canterbury, Christchurch, New Zealand November 2003. http://www.cosc.canterbury.ac.nz/research/reports/HonsReps/2003/hons_0303.pdf.

[26] A. Wallace, J. Savage, and A. Cockburn, "Rapid Visual Flow: How Fast is Too Fast?," in Proceedings of the Fifth Australasian User Interface Conference (AUIC2004), Dunedin, New Zealand, 2004, pp. 117-122.

[27] J. Williamson and R. Murray-Smith, "Dynamics and probabilistic text entry," Department of Computing Science, Glasgow University DCS Technical Report TR-2003-147, June 2003.

[28] J. Williamson and R. Murray-Smith, "Granular synthesis for display of time-varying probability densities", International Workshop on Interactive Sonification (Human Interaction with Auditory Displays), eds. A. Hunt, Th. Hermann, Bielefeld, Germany, January 2004

[29] S. Zhai, B. A. Smith, and T. Selker, " Improving Browsing Performance: A Study of Four Input Devices for Scrolling and Pointing Tasks," in INTERACT, 1997, pp. 286-292.

An Evaluation of Techniques for Browsing Photograph Collections on Small Displays

Dynal Patel[1], Gary Marsden[1], Steve Jones[2], and Matt Jones[2]

[1] Dept of Computer Science, University of Cape Town, Private Bag Rondebosch 7701, Cape Town, South Africa
{dpatel,gaz}@cs.uct.ac.za
[2] Dept of Computer Science, University of Waikato, Private Bag 3105, Hamilton, New Zealand,
{stevej,mattj}@cs.waikato.ac.nz

Abstract. In this paper we evaluate techniques for browsing photographs on small displays. We present two new interaction techniques that replace conventional scrolling and zooming controls. Via a single user action, scrolling and zooming are inter-dependently controlled with *AutoZoom* and independently controlled with *GestureZoom*. Both techniques were evaluated in a large-scale, 72-subject usability experiment alongside a conventional thumbnail grid image browser. Performance with the new techniques was at least as good as that with the standard thumbnail grid, even though none of the subjects had prior experience with such systems. In a number of cases – such as finding small groups of photos or when seeking for images containing small details – the new techniques were significantly faster than the conventional approach. In addition, *AutoZoom* and *GestureZoom* supported significantly more accurate identification of subsets of photographs. Subjects also reported lower levels of physical and cognitive effort and frustration with the new techniques in comparison to the thumbnail grid browser.

1 Introduction

The nature of photography has changed dramatically. It was once the business or pastime of a small number of individuals—experts in both the technology for capturing images and the chemistry of processing them. However, since the introduction of the Kodak Brownie a little over 100 years ago, personal photography has become increasingly affordable and pervasive. Indeed, photographic technology is now incorporated into a range of devices such as personal digital assistants (PDAs) and mobile telephones enabling photographs to be taken more quickly and cheaply than ever before. Although such devices have ever-increasing capacities to store images, their use presents users with a challenge, as the screens on which those images are browsed and viewed have become smaller.

A question that arises then, is how may a user be supported in browsing a set of photographs on such a device with limited display space? In this paper, we present two new scroll and zoom photo browsing interfaces that simplify navigation controls. Each of these interfaces utilizes two control mechanisms: one that behaves in a similar manner to a scrollbar to support scrolling and provide spatial orientation, and

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 132–143, 2004.

another that combines control over both scrolling and zooming. In the *AutoZoom* interface, this second mechanism utilizes the Speed Dependent Automatic Zooming (SDAZ) technique [8], in which scroll speed and zoom level are inter-dependent. In the *GestureZoom* interface, scrolling and zooming are controlled independently. In both interfaces distinct zooming, panning and scrolling actions are replaced with a mechanism through which control over scroll direction, scroll speed, and magnification level of the user's information space are integrated into a single action. We carried out an experimental evaluation of the two interfaces, and compared their performance to a conventional vertically-scrolled row-column thumbnail method, as is used in applications such as Apple Computer's iPhoto. Both objective and subjective quantitative measures reflect positively on the new designs.

2 Background

2.1 The Current State of Photo Browsing

To explore the sorts of features used in a digital photo organizer, Rodden and Wood studied participants' use of the "Shoebox system" [13]. This system offered advanced features such as audio and text annotation for playback, and content-based image searching. However, users took little advantage of them, emphasizing the utility of two core facilities found in many commercial photo browsers: chronological arrangement and browsable thumbnails. There are three possible reasons for these user preferences: chronological information access is natural for users as shown in the context of email [15] and personal information spaces [10]; users shy away from the computationally expensive content-based image searches, choosing to exploit the human visual system to rapidly scan and process a grid of thumbnails; and, finally, these schemes do not require user effort, like manual annotation, in organizing or pre-processing of images.

Recently researchers have proposed ways of improving on the two core facilities offered by standard commercial browsers by proposing more efficient image layout algorithms and exploiting metadata automatically added to photographs by digital cameras. Photomesa [3] is an example of a browser that uses novel layout mechanisms (quantum treemaps and bubble maps) that allows users to see as many photos as possible and maintain context. It allows users to group photographs by date, filename and directory. A PocketPC version of the system [9] has been produced but the usability evaluation did not show any improvements over the conventional approach.

PhotoTOC [11] is a browsing user interface that uses an overview and detail design. The detail view is a list of thumbnails laid out in a grid, ordered by time. The overview pane is automatically generated by an image-clustering algorithm, which clusters on the creation time and the color of the photographs. However the evaluation shows that PhotoTOC was no better, and was sometimes out performed by, Light Box (a row-column thumbnail browser which simply showed all the pictures in a flat, scrollable list, ordered by creation time).

The Calendar Browser [6] also exploits the automatically annotated timing data to structure collections of photographs into meaningful summaries. Results from a user

study show that summarized collections can lead to significant improvements in the time taken to search for an individual photograph.

While the advanced clustering techniques of the Calendar Browser and PhotoTOC browser may open up interesting ways for users to access their photograph collections, given the known preference for simple, chronological, thumbnail scrolling schemes, we were motivated to improving these within small screen contexts.

2.2 Improving Standard Scrolling Schemes

A number of researchers have been interested in improving standard scrolling schemes. Igarashi and Hinckley [8] have identified two major limitations with using traditional scrollbars. Firstly, when browsing a document, users have to shift their focus between the document and the scrollbar. They suggest that this may increase the operational time and may cause a significant attentional overhead. Secondly, they observed that in large documents, small scrollbar movements can cause a large movement of the document. This rapid rate of change can be too great for users to perceive, resulting in visual blurring and consequent user disorientation.

To counter this visual blurring, Igarashi and Hinckley proposed Speed Dependent Automatic Zooming (SDAZ). This navigation technique also alleviates other problems with conventional scrolling (e.g. attentional overhead). SDAZ unifies rate based scrolling and zooming by automatically adjusting the zoom level during scrolling to reduce the effect of rapid visual flow when a document is scrolled quickly at its normal scale. However their preliminary evaluation of SDAZ for document, map browsing and image browsing on a desktop computer produced disappointing results, with similar or worse performance than traditional methods.

Cockburn and Savage [4] carried out a substantial evaluation of their own implementations of the SDAZ document and map viewing application. Their systems used sophisticated graphical processing techniques to provide more responsive, smoother scroll and zoom animations. Their results are much more promising and show SDAZ in a new light. In their evaluation, Cockburn et al found that participants were 22% faster when using SDAZ than when using a common commercial document viewer. In map browsing, the performance benefits increased to 43%. Furthermore, workload assessments, preferences and the participant's comments all amplified the efficiency and effectiveness of the automatic zooming approach.

Both prior studies of SDAZ focused on its use on standard desktop displays, where a larger percentage of the information space is visible than is the case on small screen devices. The Palm Zire 71, for example, provides roughly 5% of the display area of a standard 15-inch laptop computer screen. The implication is that navigation may require increased user interaction for panning, zooming and scrolling when conventional navigation mechanisms are used. The experiment that we report on in the following section determines the extent to which our variations on SDAZ can ameliorate these problems for browsing photographic collections on small displays.

3 Photo Browsing on Small Displays

We developed two scroll and zoom based photo-browsing interfaces: *AutoZoom* and *GestureZoom*. In both interfaces, photographs are presented in a vertical list that is a single image wide, with a chronological ordering placing the most recent images at the top of the viewport. This organization is consistent with findings by Rodden and Wood [13], that users were satisfied with a simple chronological and folder/event based arrangement of their digital photographs, leading to more frequent browsing and reducing the effort of finding particular images. Additionally, the use of a vertical list provides methodological consistency with Igarashi and Hinckley, and Cockburn and Savage. However, we are aware that the choice of a vertical or horizontal list is language dependent (Dong et al [5]), and have designed both interfaces to allow users to configure scrolling direction

For the *AutoZoom* interface, the SDAZ variant is operated by vertical dragging actions with the pointing device. These actions control the rate at which images scroll through the viewport, the image size (zoom level) and the scroll direction. The vertical centre of the viewport acts as the threshold for direction change—dragging above the centre moves the images downwards and vice versa. Image size is inversely proportional to the distance of the pointer from the vertical centre, and changes dynamically as the pointer moves either away from or towards the centre (see Figure 1). Images are not reduced beyond a minimum (user specified) size threshold. Once this threshold is reached, an acceleration function maps further increases in drag distance proportionally to scroll speed.

The perceived effect to the user, then, is that as they increase their scrolling speed, the photo images get smaller and smaller, zooming out to get an overview, reducing the effects of visual blur. When the user completes an action by releasing the pointing device, the images are smoothly animated back to their normal size at the current location in the list.

For the *GestureZoom* interface, vertical drag operations control scroll speed and direction as with the *AutoZoom* interface, but do not control image size (zoom level). Zoom level is controlled by horizontal movement of the pointing device away from the horizontal centre of the viewport to the right-hand or left-hand side of the display. Image size is inversely proportional to the horizontal drag distance.

Figure 2 (a) shows a pointer position – indicated by the cross – leading to a moderate scroll speed with small image reduction: the user is dragging below and slightly to the right of the viewport centre. In 2 (b), the user has dragged the pointer to the right-hand corner of the display, producing the maximum scrolling speed and the minimum image size. Returning the pointer to the centre of the viewport returns the images to the full size. As with *AutoZoom*, when the user releases the pointer (e.g. removing the stylus from the screen), the images smoothly animate back to their normal size.

The scrollbar has the same appearance and behaviour in the two interfaces—as the user begins to drag the slider the image thumbnails are immediately reduced to their minimum size and normal scrolling follows. At the end of a scrolling operation the images are expanded to their normal size. Hence, the scrollbar can be used for quickly gaining an overview of the image set, allowing users to find an approximate location in the set of photographs. Our approaches extend the original SDAZ implementations

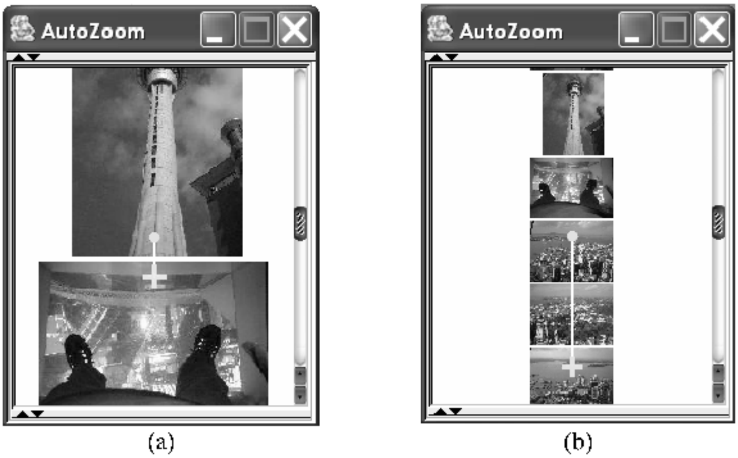


Fig. 1. *AutoZoom* interface: as cursor is dragged away from centre, scroll speed/image size change. (a) moderate speed, images slightly reduced; (b) faster speed and smaller size (Cross added for clarity)

Fig. 2. *GestureZoom* interface: (a) moderate scroll speed and small image reduction; (b) maximum speed and minimum size. (Cross added for clarity)

in a number of ways ([4],[8]). For instance, our algorithms have been developed to allow support a range of small screen sizes and input devices (see Section 7); they present a simple control feedback (the vertical line); and, the navigation direction can be set to either vertical or horizontal to support language differences.

A further browser called the *DiscreteZoom* browser (see Figure 3) was implemented for the purposes of comparative evaluation. It is a thumbnail browser that presents photographs in row and column scrollable list ordered by creation time. Users can click/tap on the desired photo to view an enlarged version. The selected photo is animated to fill the screen. Similarly users can click/tap on the enlarged photo to return to the thumbnail view. This browser reflects the features found in

(a) (b)

Fig. 3. *DiscreteZoom* browser: (a) the thumbnail view; and, (b) the enlarged view

popular commercial browsers such as Apple iPhoto or ACDSee Picture Viewer [1],[2].

4 Evaluation

4.1 Hypotheses

The objective of the experiment was to compare user performance and subjective preferences with each of the three photo navigation techniques. Our hypotheses were as follows:

1. both *AutoZoom* and *GestureZoom* support faster navigation to target photographs than *DiscreteZoom*;
2. both *AutoZoom* and *GestureZoom* support more accurate identification of target photographs than *DiscreteZoom*;
3. subjective task load levels are lower for both *AutoZoom* and *GestureZoom* than *DiscreteZoom*.

4.2 Subjects

Seventy-two subjects (38 male and 34 female) took part in the experiment. Sixty-one subjects were students (either postgraduate or undergraduate), 6 were lecturers and 5 were software developers. 45 of the subjects had previously used photo management software, but only 5 on a small screen device. None of the subjects had used SDAZ interfaces. 70 participants described themselves as casual photographers (i.e. occasionally take photographs). Two participants described themselves as professional photographers (e.g. take photos for magazines or weddings).

4.3 Method

A repeated measure factorial design was employed. Subjects were randomly allocated to one of three groups, each containing 24 subjects. Each group used only one of the three interface designs to complete photo navigation tasks.

The independent variables were as follows:

- Interface. Between-subjects variable with three levels: *AutoZoom*, *GestureZoom* and *DiscreteZoom*;
- Task type. Tasks-types were based on those identified by as key to photo-browsing [13]. The type was within-subjects variable with three levels: *Event* (subjects searched for a set of photos relating to a particular well-defined event, e.g., "locate the Motor Rally"); *Single* (subjects searched for an individual photo containing a specified *Feature,* e.g., "Find this image of the Sky Tower"); and, *Property* (subjects searched for a set of photos taken at different events, but all sharing a property, such as all the photos containing an specific object, e.g., "Count all the photos that contain an hot-air ballon");
- Navigation distance. For *Event* and *Single* task types only. Within-subjects variable with two levels: short and long. Short distances were no more than half the length of the photograph list, and long distances were always more than half the length.

Events could be small (3 or fewer photos), or large (more than 3 photos – Figures 1 & 2, then, contain large events). A photograph *Feature* could also be small or large: a small feature was one that was $1/8^{th}$ or less of the total image size (e.g. a small child in a forest scene), while a large feature was one taking up more than $1/8^{th}$ of the image (e.g. a skyscraper).

Each subject completed a total of 27 experimental tasks, using one of the interfaces. For the *Event* task type they completed 3 tasks for each of the 4 navigation distance/event size combinations. For the *Single* task type they completed 3 tasks for each of the 4 navigation distance/feature size combinations. For the *Property* task type they completed 3 tasks (requiring the user to find 16, 30 and 120 images respectively).

Presentation order of the tasks was counterbalanced to minimize learning effects.

4.4 Experimental Measures

For each task the software automatically recorded a range of events including: time to complete task, distinct scrollbar operations and distinct zoom operations.

For *Property* tasks there was a target number of photos (A); in completing the task, a user found a number of images (C). Accuracy was then calculated as:

$$Accuracy = 100\left(1 - \frac{|A - C|}{A}\right)$$

Finally, we collected subjective responses about the workload required to complete tasks, as measured by the NASA task load index [7]. Responses were on a scale of 1 to 5, with lower values reflecting lower task loads. In all cases, the statistical data was subjected to significance testing using the analysis of variance method (ANOVA).

4.5 Procedure

On arrival, subjects were asked to read a summary of the experiment and provide consent to continue if they were in agreement. They then completed a profile questionnaire and were given 15 minutes to familiarize themselves with the set of photographs to be used in the experiment. At the end of this time, they were required to read instructions that provided a detailed description of each task type and also explained the operation of their assigned interface.

The operation of the interface was then demonstrated, and the subjects were given 10 minutes to explore the operation of the software for themselves. Following this they were given a set of training tasks of the same form as the experimental tasks. Subjects were encouraged to ask questions throughout the training period. Once the training tasks were completed subjects could take a short break, before commencing the experimental tasks. One aim of the training session was to allow users to familiarize themselves with the image set so that any learning effects during the experimental tasks would be reduced.

Subjects controlled progress of the experimental session via an on-screen dialog that allowed them to initiate a task, displayed task instructions, and allowed them to indicate completion of a task. At the start of every task, the viewport was reset to the show the beginning of the image list.

Event tasks were described textually. An event was found by selecting any one of the photographs within the event. For *Single* tasks, subjects were shown the target photograph and its corresponding caption. For both *Event* and *Single* tasks, users were prompted by the system to retry if their selection was incorrect; they were able to attempt the task as many times as they needed.

For *Property* tasks, subjects were required to count the number of photographs that shared a common property. They were given a field into which to enter the number. On completion of all the tasks subjects were requested to fill-out a questionnaire that captured their subjective views of the software and workload estimates via a NASA Task Workload Index.

4.6 Materials

The experiment was carried out on a desktop computer with a 1.7GHz processor, 1GB of RAM, and running Microsoft Windows XP. The viewport size for all three interfaces was set to 240x340 pixels to simulate the display of the HP h5550 Pocket PC. Users used a mouse as a stylus surrogate.

A single set of 300 of photographs was used in the experiment, providing a consistent set of stimuli across all tasks, subjects and conditions. The photographs were typical tourist type images – beach and mountain scenes; individuals and groups in sightseeing locations; and significant events, such as holiday periods – gathered over a 6 month visit to New Zealand by one of the authors.

5 Results

5.1 Locating *Events*

AutoZoom and *GestureZoom* interfaces were significantly faster than the *DiscreteZoom* interface when searching for small events ($F(2,69) = 5.0597$, $p=0.00890$), with means of 26.0 seconds, 29.4 seconds and 45.5 seconds, respectively. Over all *Event* tasks, though, interface type had no significant effect on task completion time ($F(2,69)=1.2848$, $p=0.28323$).

Regardless of interface, subjects took significantly *longer* to locate events which were a short distance away ($F(1,69) = 8.9667$, $p=0.00381$), with short and long distance means of 33.25 and 23.46, respectively. At long navigation distances, large events were found significantly faster than the small events ($F(1,69)=6.5946$, $p=0.01240$), with mean search times of 14.04 seconds and 32.88 seconds. For short navigation distances, event size had no significant interaction with the time to locate an event, with search means of 34.38 seconds (large events) and 32.12 seconds (small events).

5.2 Locating *Single* Photographs

AutoZoom was significantly faster at finding single photographs than *DiscreteZoom* at long navigation distances ($F(1,46) = 9.5749$, $p=0.00335$) with means of 28.90 seconds and 44.06 seconds, respectively. Both *Autozoom* and *GestureZoom* were significantly faster than *DiscreteZoom* when searching for images with small features ($F(2,69) = 3.1596$, $p = 0.04865$) with means of 39.15 seconds, 34.52 seconds and 48.69 seconds, respectively. Over all *Single* tasks, though, interface type had no significant effect on task completion times ($F(2,69)=0.79012$, $p=0.45785$).

Regardless of interface, subjects took significantly less time to locate single images that were a short distance away ($F(1,69)=11.330$, $p=0.00125$), with short and long distance means of 26.85 seconds and 34.98 seconds respectively. Also, images with smaller features took significantly longer to detect than those with larger ones ($F(1, 69)=61.446$, $p=.00000$), with small and large means of 40.79 seconds and 21.04 seconds respectively.

5.3 Locating Photographs with a *Property*

AutoZoom and *GestureZoom* were significantly more accurate than *DiscreteZoom* ($F(2,69)=14.614$, $p=0.0001$), with mean accuracy rates of 92.38%, 89.98% and 76.15%, respectively. Over all *Property* tasks, interface type had no significant effect on task completion time ($F(2,69)=1.5150$, $p=0.22704$).

5.4 Subjective Preference

There was a significant difference between the mean task load ratings for the three interfaces ($F(2,69) = 6.0275$, $p=0.00387$): the mean rating for *DiscreteZoom* was 3.01; for *Autozoom* it was 2.31; and, for *GestureZoom*, 2.53.

Looking at the individual factors measured by the task load index, subjects found both new interfaces significantly less frustrating than the *DiscreteZoom* interface $F(2,69) = 7.9593$, $p= 0.00078$). Furthermore the mental workload ($F(1,46) = 8.4033$, $p = 0.00572$) and effort ($F(1,46) = 7.9310$, 0.00713) were significantly lower for the for the *AutoZoom* interface than the *DiscreteZoom* interface.

6 Discussion

Considering the results in the light of the three hypotheses noted in Section 4.1.

1. **Both *AutoZoom* and *GestureZoom* support faster navigation to target photographs than *DiscreteZoom*.** The results indicate the new techniques performed as well and in some cases better than *DiscreteZoom*. More specifically, both new interfaces were significantly faster when finding *Single* photos containing small-sized features as well as detecting *Events* consisting of a small number of photos. *AutoZoom* was also significantly faster than the *DiscreteZoom* interface at locating *Single* images at long navigation distances.
2. **Both *AutoZoom* and *GestureZoom* support more accurate identification of target photographs than *DiscreteZoom*.** The new techniques were significantly more accurate when finding a set of photographs that fit a given description.
3. **Subjective task load levels will be lower for both *AutoZoom* and *GestureZoom* than *DiscreteZoom***. The results of the task load calculations show that subjects perceived the new systems to be significantly less onerous than the *DiscreteZoom* browser.

It is worth remembering that none of the subjects had previous experience of SDAZ-type interfaces while all would be familiar with the conventional scrolling approach of *DiscreteZoom*. It is encouraging, then, to see such consistently good performance with the new schemes after minimal training. During task completion, the average amount of time spent operating the zoom/scroll control with the new interfaces was 22.5 seconds; this is nearly four times the duration spent using the scrollbar (5.9s). We are satisfied, then, that the benefits provided by the new interfaces come from the integration of scrolling and zooming.

Small features in an image, small groups of photographs and individual, target photos are more easily overlooked with *DiscreteZoom*, as they scroll past at thumbnail size; the explicit zoom-in/zoom-out operations needed to check individual image contents also contributes to the slower performances. Such problems with grid-based thumbnail browsing have been recognized by others who suggest, for example, processing the images to present only the salient details [14]. *AutoZoom's* better performance at finding *Single* images at long navigation distances suggests that these sorts of technique may be of greater benefit for very large sets of image.

7 Future Work

This experiment was simulated on a desktop computer as at the time the software was written, PDAs such as the HP Pocket PC did not have sufficient processing power and

memory to run such applications. The apparatus has allowed us to gain very useful insights into the relative benefits of browsing schemes. We have now ported the code to a mobile environment, achieving responsive, smooth animation.

While the approaches have been implemented to accommodate a device using a pointer (e.g. a stylus), they can be extended for use with other interaction devices. For example, *AutoZoom* could be used with physical dial-type wheels as seen on the iPod or the *smartPad* proposed by Rekimoto for use in mobile phones [12], providing one-handed interaction. Meanwhile, joystick-type mechanisms may permit the use of *GestureZoom* schemes.

8 Conclusions

Our work provides evidence that small screen photo browsing may be improved with interaction schemes that integrate scrolling and zooming. As camera enabled mobile devices become more common, and picture taking and sharing more prevalent, it will become increasingly important to manage photograph collections using a small screen and input devices such as a stylus. We believe that the work presented here forms a good foundation for future generations of this software.

Acknowledgements. This work was supported by grants from the University of Waikato, the University of Cape Town and from Microsoft Corporation. Stephen Brewster and Richard Schroder provided helpful inspiration with respect to *GestureZoom.* We are also grateful for the comments of the anonymous reviewers.

References

1. ACDSee (2001). http://www.acdsystems.com/english/products/acdsee/acdsee-node.htm
2. Apple iPhoto. (2002). http://www.apple.com/iphoto
3. Bederson, B., (2001). Photomesa: A zoombable image browser using quantum treemaps and bubblemaps, in *Proceedings UIST 2001*, 71-80, ACM Press.
4. Cockburn, A. & Savage, J. (2003). Comparing Speed-Dependent Automatic Zooming with Tradional Scroll, Pan, and Zoom Methods, *People and Computers XVII: British Computer Society Conference on Human Computer Interaction. Bath, England. 2003*, 87-102.
5. Dong, Jianming & Salvendy, G. (1999). Designing menus for the Chinese Population: Horizontal or Vertical? *Behaviour and Information Technology* Vol. 18, 467-471.
6. Graham, A., Garcia-Molina, H., Paepcke, A & Winograd, T. (2002). Time as essence for photo browsing through personal digital libraries. In *Proc. ACM/IEEE JCDL (2002)*, 326–335. ACM Press.
7. Hart, S. & Staveland, L. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research, in P. Hancock & N. Meshkati, *eds*, *Human Mental Workload*, 139–183, Elsevier Science.
8. Igarashi, T. & Hinckley, K. (2000). Speed-dependent automatic zooming for browsing large documents, *Proc. UIST 2000*, 139-148, ACM Press.
9. Khella, A. & Bederson, B. (2003). A Zooming Image Browser for the Pocket PC. University of Maryland, Technical Report: www.cs.umd.edu/~akhella/pocketphotomesa.pdf

10. Krishnan, A. (2003*). Interactive Time-Centered Workspace Visualization*. MCMS Thesis Department of Computer Science, University of Waikato, Hamilton, New Zealand. Call no. TK7882.I6K75.
11. Platt, J C., Czerwinski, M. & Field, B. (2002). PhotoTOC: Automatic Clustering for Browsing Personal Photographs, *Microsoft Research Technical Report MSR-TR-2002*.
12. Rekimoto, J., Oba, H. & Ishizawa, T. (2003). SmartPad: a finger-sensing keypad for mobile interaction, *Proc. of Human Factors in Computing Systems, CHI 2003*, 850–851. ACM Press.
13. Rodden, K & Wood, K. (2003). How do people manage their digital photographs? *In Proc. of Human Factors in Computing Systems, CHI 2003*, 409-416, ACM Press.
14. Wang, M-Y, Xie, X, Ma, W-Y & Zhang, H-J. (2003). MobiPicture - Browsing Pictures on Mobile Devices, *Proceedings 11th ACM International Conference on Multimedia*, 106-107, Berkeley, CA, USA, Nov. 2003.
15. Whittaker, S. & Sidner, C. (1996). Email overload: exploring personal information management of email, in *Proceedings of Human Factors in Computing Systems, CHI 1996*, 276–283. ACM Press.

Variability in Wrist-Tilt Accelerometer Based Gesture Interfaces

Andrew Crossan[1] and Roderick Murray-Smith[1,2]

[1] Hamilton Institute, National University of Ireland, Maynooth, Co. Kildare Ireland.
andrew.crossan@may.ie
http://www.hamilton.ie/andy

[2] Department of Computing Science, University of Glasgow, Glasgow, Scotland.
rod@dcs.gla.ac.uk
http://www.dcs.gla.ac.uk/~rod

Abstract. In this paper we describe a study that examines human performance in a tilt control targeting task on a PDA. A three-degree of freedom accelerometer attached to the base of the PDA allows users to navigate to the targets by tilting their wrist in different directions. Post hoc analysis of performance data has been used to classify the ease of targeting and variability of movement in the different directions. The results show that there is an increase in variability of motions upwards from the centre, compared to downwards motions. Also the variability in the x axis component of the motion was greater than that in the y axis. This information can be used to guide designers as to the ease of various relative motions, and can be used to reshape the dynamics of the interaction to make each direction equally easy to achieve.

1 Introduction

Mobile devices are now widely used for a variety of everyday tasks. However, due to the requirement for a small screen, interacting with these devices often proves to be difficult. On-screen buttons are generally closely grouped together making interactions slow and error prone. This is particularly the case in a mobile context where the user's visual attention may be required elsewhere.

Generally, interaction with these devices has taken the form of discrete messages passed between the user and device. The user will click a button or select a menu item, and the device will supply feedback. This method can be slow and frustrating particularly in situations requiring many button clicks such as typing with an on-screen keyboard.

The development of new interaction techniques and sensors provide more opportunity for a more continuous form of interaction, allowing closed loop interaction between device and the user's motions. In this instance, all of the user's movements affect the interpretation of the interaction and the device can continually change the feedback supplied to the user accordingly. Gesture input is one form of continuous interaction that has been underused in interaction with current systems. Text entry is

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 144–155, 2004.

the one major exception to this where gesturing with a stylus is often used for inputting text to a PDA. In this case, it is used to provide a quick, more natural alternative to a screen-based keyboard where the keys may be required to be small and are tightly packed together leading to high error rates. Pirhonen, Brewster and Holguin [6] demonstrate an example of gesturing as an input technique for controlling a PDA based MP3 player. These interactions are designed to be intuitive for the task performed. Pirhonen, Brewster & Holguin were able to demonstrate significant usability benefits with the gesture interface over the standard interface, with users indicating that the gesture system required a lower workload to perform the task.

Recent studies have examined the possibility of using accelerometers attached to a mobile device to provide input. Advantages over most stylus based gesture systems are that they offer the possibility of one handed, screen free gesture control. They are often suggested as useful for continually monitoring background acceleration and providing context information for the current task. The components required for inertial input are also cheap to manufacture. (ca. $2 a device for mass production).

Accelerometers allow a user to input data and commands by tilting the device. Hinkley *et al.* [2] present a study that demonstrates a tilt-based gesture system for scrolling and automatic screen orientation of a PDA. Through user testing, they were able to provide a system that performed screen orientation and scrolling in a manner that was useful and predictable to the user. This study demonstrates the potential for tilt-based gestures to provide a fast, natural method for interaction.

Rekimoto [7] explores the possibility of using tilt input to navigate menus and scroll large documents and maps. The prototype system described allowed users to select items in pie menus although no formal evaluation was carried out.

Williamson and Murray-Smith [9] have developed the Hex system for inputting text on a PDA with accelerometer. This system allows the user to select letters by tilting the PDA to navigate a cursor through a series of tiled hexagons. Through use of a language model, they were able to adjust the feedback given to the user such that probable sequences of characters were easier to perform than non-probable sequences. TiltType presented by Partridge *et al.* [5] is similarly a tilt based text entry method where characters are selected by a combination of button clicks and the orientation of the device. The inertial control allows TiltType to be used on devices with extremely small screens such as a watch.

2 Targeting Tasks

There is a large body of literature studying targeting tasks using many different input devices. Most common are Fitts' Law based studies where users are required to continuously move between two targets (an overview can be found in [3]). Timing and error rates can be gathered for different target widths and separations allowing the experimenter to calculate the comparative difficulty of the task. Most studies work with univariate targets by setting narrow target widths while allowing effectively infinite target heights. Accot and Zhai [1] describe a study that extends Fitts Law to

take account of two-dimensional targets. Their experiment was used to select a model that provides the Fitts' Law index of difficulty for two-dimensional targeting.

MacKenzie *et al.* [4] describe methods that are based on the variability in movement rather than error rates. They suggest task metrics suitable for measuring movement variability including slip off errors, mean distance from the task axis, movement variability perpendicular to the task axis, and orthogonal direction changes.

This paper is concerned with gesturing using wrist tilt motions. With all gesturing systems, there will be a degree of variability in the gesture, and therefore uncertainty about the gesture performed. This study examines the variability in movement for short gestures in eight directions. The gestures require users to move a cursor between a series of pairs of points by tilting their wrists. The study hoped to determine areas of difficulties at the limits of comfortable movement in different tilt directions. Both error rate and variability metrics are considered. Speed and accuracy of targeting in different directions is also examined.

3 Experimental Method

3.1 Equipment

The experiment was carried out with an HP 5450 PDA with the Xsens P3C 3 degree of freedom linear acceleration sensor attached to the serial port (shown in Figure 1). Its effect on the balance of the device is negligible (its weight is 10.35g). The accelerometer was used to detect tilt magnitude around the x and y axis of the mobile device, sampling at a rate of 35 samples per second.

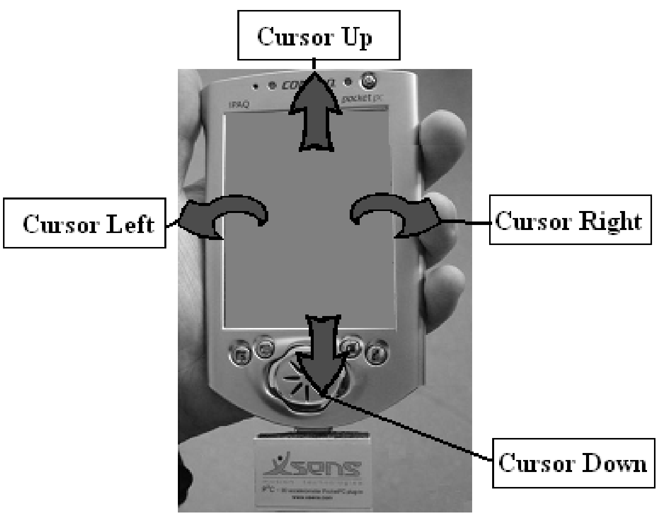


Fig. 1. PDA with XSens accelerometer attached at the base. The user would move the cursor by tilting the device in the directions shown.

3.2 Task

The experimental environment used is shown in Fig. 2. Nine circular targets of radius 15 pixels were placed throughout the environment. One target was placed at the centre of the screen, and eight were spaced at 45-degree angles around the circumference of a circle centred on the initial target such that the radius of this circle was 100 pixels. The gain on the cursor movement was set such that this distance corresponded to a tilt of approximately 48 degrees in the x direction and approximately 36 degrees in the y direction. The difference in these values correspond to a scaling due to screen size such that the same tilt magnitude is required to move to each of the edges of the screen (for a screen of width 240 pixels and height 320 pixels). Due to the different x-y cursor gains, the results section considers comparisons made between targets in opposite directions only.

These values provided a wide range of tilts while still allowing the user to easily view his or her interaction on the screen. A pilot study suggested that screen contrast became an issue with larger tilts in the y direction. The cursor gain was deliberately set to a low value such that large tilts would be required to complete the task and the limits of the movement would therefore be explored.

The task given to participants was to select the highlighted target (in Fig. 2 the top centre target is shown to be highlighted). The cursor was controlled by a linear position control mechanism, mapping rotation of the device to movement of the cursor. The device accelerometer was calibrated such that the starting position of the device corresponded to the centre position on the screen. This calibration occurred at the start of each trial. To move the cursor in the x direction, the device was titled left or right, and to move the cursor in the y direction, the device was tilted up or down (shown in Fig. 1). Distance of the cursor from the centre position was directly mapped to angle of rotation from the rest position. Therefore, double the rotation angle of the device would lead to the cursor being twice as far from the central position. Since a position control mechanism was employed, if the user held the devices still at any orientation, then the cursor would remain still on the screen.

Users held the device in their dominant hand and were instructed to sit in a comfortable position with the device held such that they could easily see the screen. In practice, all participants sat with the device slightly tilted towards them and leaning forwards slightly over the device.

Selection required the user to hover the cursor over the target for 1.5 seconds. If the cursor slipped off the target before the selection was complete, the target timer was reset and the user was again required to move onto and hover over the target for the full one and a half seconds. Once successful selection of a target was complete, a different target was then highlighted. The sequence of targets was chosen such that highlighted targets alternated between any of the outside target and the centre target. This ensured that all movement was either from the central target to an outer target, or from an outer target to the central target. This sequence was chosen to ensure that the path distance to the next target was always kept constant and that the angle to the next target was restricted to the eight equally spaced angles chosen.

Two competing factors affected the chosen target size. As the trajectory rather than the targeting was the main measurement for the task, the targets needed to be big enough to allow easy targeting. However, to maintain similar path length between starting position and target position, the targets could not be made to be too large. A diameter of 15 pixels was eventually chosen empirically. A bar at the top of the screen (shown in Fig. 2) indicates the time the user has spent over the target. When the bar reaches the right of the screen, target selection has been completed.

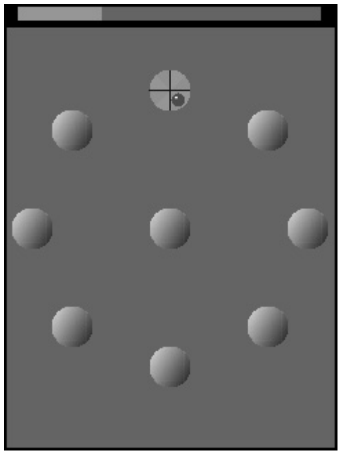


Fig. 2. The experimental environment used for the study. The top centre target is the highlighted target. The cursor is the smaller circle within this target.

All participants took part in three experimental sessions with an hour break between each for recovery. The first session was used to train users in the task. The second and third sessions were eventually used when analysing the movement characteristics of different participants. The sessions were designed to be short to minimise user fatigue. No session lasted over five minutes.

3.3 Participants

Twelve participants took part in the training then the two experimental sessions. Their ages ranged from 23 to just under 40 and eleven were male. Two had previous experience with accelerometers and mobile devices, but none had experience with the cursor control mechanism described above. Ten participants were right handed and two were left handed, and all used their dominant hand for this study. The effect of this factor is considered in the next section.

3.4 Hand Used to Tilt the Device

The hand used by the participant to tilt the device is an important factor when it comes to analysing the results. It is not uniformly easy to tilt the wrist in all direc-

tions, and the degree of tilt possible from a given starting position will be different in different directions. For right-handed users, to move the cursor to the right of the screen will require the wrist to be tilted such that the palm of the hand moves towards the wrist. For a left-handed participant moving the palm of the hand towards the arm will move the cursor to the left of the screen. This reversal is only true in the one axis of the wrist. Since this study is examining the restrictions placed by the body on wrist tilting interfaces, when analysing the results we must take into account the hand used by the participant during the study. The correction made for left handed participants is to switch the results obtained for targets on the left with the corresponding target on the right such that the top-left target switches with the top-right target, the rightmost target switches the leftmost target, and the bottom-right target switches with the bottom-left target.

3.5 Measured Factors

Slip Off Errors

A slip off error occurs whenever the user moves off the current target before selection. By measuring slip off errors, we can determine how difficult the targeting task was in the different directions and make comparisons. A slip off error and recovery is demonstrated in Fig. 3.

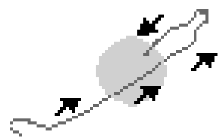


Fig. 3. Cursor trace showing a slip off error and recovery.

Trajectory Analysis

It is important to consider the ease of movement in different directions when creating gestures. This is particularly the case for applications involving rotation of the wrist where some directions may be more difficult to tilt in than others. Data was separated into different directions of movement (to the different targets) and analysis was carried out to look for paths that resulted in a high degree of variability from the ideal (direct) path to the different targets. The measure of variability used was the distance travelled when moving between targets. Moving from the central target to the edge of any of the outer targets in a straight line was 85 pixels in length. Excess path length was therefore classified as the distance travelled above this minimum.

Time to Target

This factor will measure the time taken for the user in moving onto the target. It does not include the time required to hover over the target to perform the selection.

Unintentional Movements by the User
This factor measured the noise generated by a user when holding the device still at different angles. The user hovered over a target for one and a half seconds. By analysing the middle second of this data, it is possible to estimate this noise value. This was then compared to noise generated by the sensor on a fix surface.

4 Results and Discussion

4.1 Slip Off Errors

The mean numbers of slip off errors for all users in all directions are shown in Fig. 4. These data are shown as the mean number of times that the user slipped off each target during the experiment. Each target had to be selected 12 times by each user.

The mean number of slip offs for each is relatively small when compared to the variability in the data. These data suggest that users found moving to the lower targets easier than moving to the upper targets. For the top centre target, one in four of attempts to select a target resulted in slipping off the target. This is reduced to approximately one in six attempts when targeting the bottom centre target. It must be noted that there is a high level of variability in the data.

In total, there were 309 slip off errors out of 1152 targeting attempts. This number is high compared to targeting studies with other devices. This could indicate the difficulty of the task, but could also be due to the fact that users were required to hover over a target rather than click on it.

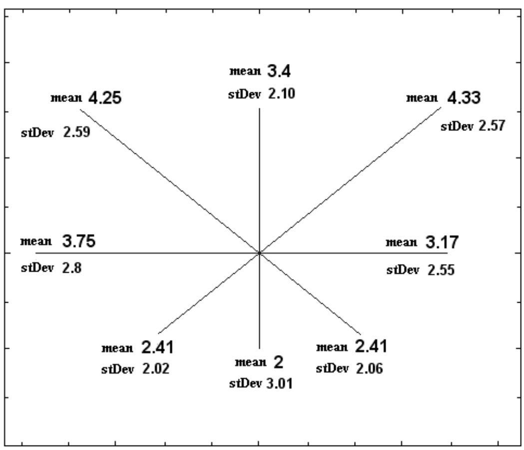


Fig. 4. Mean slip off errors for each user for all targets. Each user had 12 attempts to acquire each target.

With the low cursor gain used in this study, a lower number of slip off errors may be expected since a comparatively large tilt is allowed before slipping of the target. However, in this study, users are being asked to make large movements that required

them to rotate their wrist to the limit of movement. Future studies should investigate a higher gain that allows targeting with more comfortable ranges of movement.

4.2 Trajectory Analysis

From analysis of the cursor traces, and post hoc discussion with participants, it became clear that the mapping from wrist orientation to cursor position was confusing in a small number of cases. One user in particular had expectation of the opposite mapping. The trajectory data was initially analysed to detect cases where this occurred. These cases were defined as cases where the user initially moved at least one cursor radius (15 pixels) away from the target from the start position. Three examples of such trajectories are shown in Fig. 5. There were 30 out of 1152 such targeting attempts, which were spread over both experimental sessions. One user who expressed a strong opinion for the opposite mapping was responsible for 16 of these trajectories. Although, targeting was achieved without this confusion in the vast majority of the cases, these results suggest that the natural mapping is not as strong as in a similar position-control device, such as a mouse. Unlike when using the mouse, users must map a rotation to a cursor translation. More than one sensible mapping exists and different users may have different preconceptions of this mapping, making it more difficult to learn the opposite mapping. In this study, the cursor could be thought of as a marble attached to a piece of elastic. If you tilt one side of the device downwards from the start position, the cursor will move towards that side. One alternative model would be to think of the cursor as a bubble in liquid beneath the screen. This would correspond to the opposite mapping where tilting one area of the screen upwards would cause the cursor to move towards that area of the screen. These results show, however, that most users were comfortable with the mapping described.

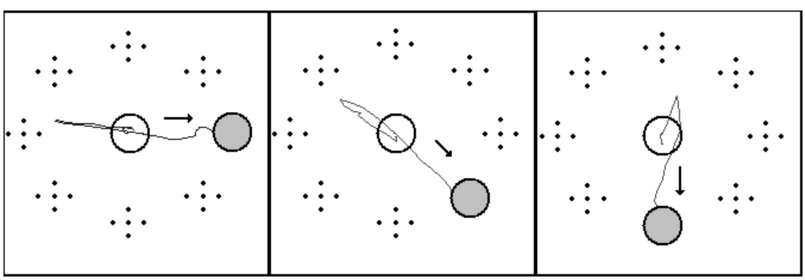


Fig. 5. Three examples of the user initially moving in the wrong direction from a centre target to the highlighted outside target.

Directional errors in movements caused by the user mistakenly moving the control device in the wrong direction have been noted by Sheridan [8]. For the errors discovered, the user consistently moved in the opposite direction from the new target. This strongly suggests confusion with the mapping rather than false anticipation of the next target. As these trajectories are most likely an artefact of confusion with the

mapping rather than difficultly in the task, they are excluded from the final analysis of the trajectory lengths.

Fig. 6 shows the mean excess difference travelled by all users when travelling to the different targets. It can be seen from this figure that some directions are easier to travel in than others. Generally, the data indicates that users found selecting targets in the bottom half of the screen easier than in the top half. These results similarly suggest that lower targets are easier to select than the higher targets. Although slip off errors will have an effect, this can be considered to be minimal as the user's movements will be comparatively small when close to the target attempting to remain in the target area. Again, the high level of variability in the data should be noted, particularly when comparing variability for the targets in the upper area of the screen with those in the lower half of the screen.

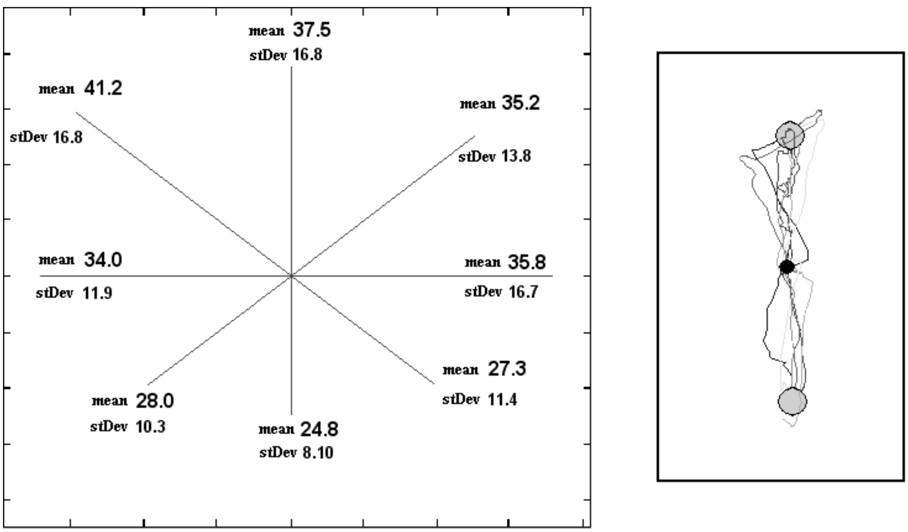


Fig. 6. (Left) Mean excess distance travelled to each target in pixels. The distance travelled during target selection is not included in this measurement. (Right) Cursor trace of one user moving to the top and bottom targets six times each during one session demonstrating variability during an individual trial.

The right of Fig. 6 displays six trajectories for a typical user targeting the top centre and bottom centre targets in the same experimental session. The variability displayed can be used to explain the longer path length noted when moving to the upper targets. This can be explained by the dynamics of the arm. For a posture where the user holds the device and looks at the screen, it is difficult and uncomfortable to rotate the hand such that the palm faces upwards and the screen is still at the appropriate rotation. There is a far greater range of movement when rotating the palm downwards.

4.3 Timing Data

The mean time to target data for all users is displayed in Fig. 7. The differences in time between the upper and lower targets are small in this instance and may be explained by the larger number of slip offs in the upward direction. This suggests that time to target is approximately uniform in all directions for wrist tilt applications.

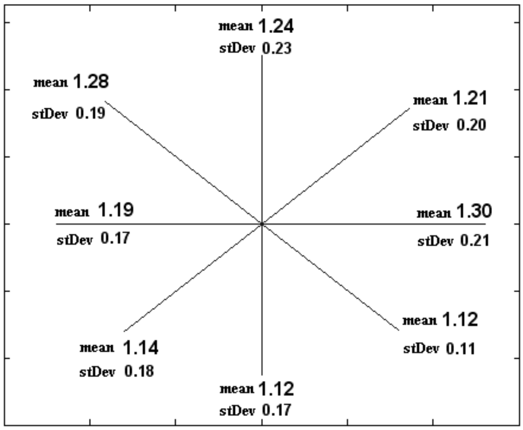


Fig. 7. Targeting time in second for each of the outer targets

4.4 Unintentional Movement by the User

Unintentional tilts generated by the user while hovering over a target were measured in the x and y directions. A measure of the variability was given by taking the standard deviation of the mean change in tilt value during one sample point for each individual target. Only the middle second of data during the target selection was considered to allow for the user moving onto the target and moving in anticipation of the next target. For illustrative purposes, sensor readings have been converted to approximate angle in degrees.

These values are shown to be consistent for all targets in x and in y. Although, differences are small, the variability in the y direction seems to be consistently smaller than the equivalent in the x direction. This could be due to the targets being smaller in the y direction due to the higher y gain. However, since the target radius in each direction would allow for a rotation of approximately 7.2 degrees in the x direction and 5.4 degrees in the y direction which is significantly higher than the variability values recorded. One other possibility to be considered is the positioning of the accelerometer. As the accelerometer is placed at the centre of the base of the device, it is at the centre of rotation in the x direction but offset in the y direction. This means that for the same tilt in x and y, the extra leverage due to the displacement of the accelerometer in y will lead to higher accelerations in that direction. If this were the cause, the opposite effect would have been expected since smaller tilts in the y direction would have moved the accelerometer a larger distance.

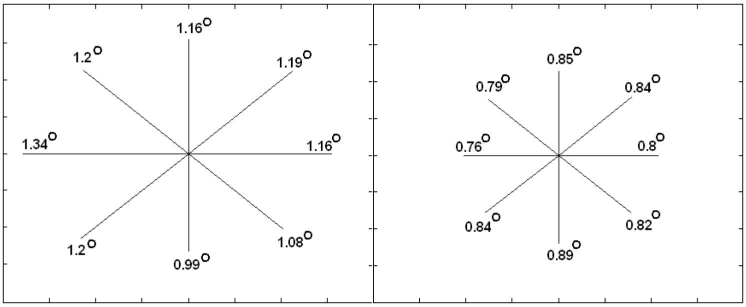


Fig. 8. Approximate variability in degrees when hovering over the target in the direction indicated. (Left) is X variability. (Right) is Y variability.

When the device was flat and at rest on a solid surface, the device generated the equivalent of 0.26 degrees tilt in x and 0.24 degrees tilt in y. It can be seen from Fig. 8 that these values are far smaller than the measured values for the device when held by a user at different angles.

By constantly monitoring the variability of the sensor readings, it is therefore possible to detect when the user is holding the device in a controlled fashion, and when it is resting on a surface. This provides similar functionality to that proposed by Hinkley *et al.* [1] but using accelerometer data rather than an extra touch sensor. This context information would provide programs running with information about the state of use of the device that can be used to modify its behaviour.

5 Conclusions and Future Work

This paper has examined the variability in movement in different directions for short wrist-based target acquisition with visual feedback. The results demonstrate that the direction of cursor movement affects the performance of the user in a tilting task. With the marble control metaphor described, users displayed more variability and lower performance when moving to targets in the upper half of the screen compared to targets in the lower half of the screen. No time difference was detected when moving to the upper or lower targets. The results suggest a high level of variability in the movements. It should be noted, however, that the system described in this study was not designed to produce optimal targeting results but explore variability in motion. Performance would be expected to improve with a higher cursor gain and different selection mechanism. This information can guide interface designers, as to the relative difficulty of different tilt-motions.

The ease of use of the mouse has demonstrated how a non-linear control display gain can provide a natural mechanism for interaction. Our future work will look at inverting our model for wrist-based tilting to enable us to achieve uniformly easy tilting behaviour in all directions. There is the potential in tilt-based interfaces to compensate for different levels of variance in different directions by adapting the dynamics of the cursor depending on the state and velocity vector - the handling

qualities would be more damped in regions of higher variability. The trajectories will be further analysed to examine the possibility of using the individual user variations and movement characteristics to identify that user.

Future studies will initially examine wrist tilt cursor control with higher gain levels and eventually lead to developing interactive systems that provide changing dynamics to aid the user's movements, and reduce variability. These methods will also be applied to coping with disturbance, particularly for interaction in a mobile context.

Acknowledgements. The authors gratefully acknowledge the support of SFI project *Continuous Gestural Interaction with Mobile devices*, Science Foundation Ireland grant 00/PI.1/C067, the *Multi-Agent Control* Research Training Network - EC TMR grant HPRN-CT-1999-00107, and EPSRC grant *Audioclouds: three-dimensional auditory and gestural interfaces for mobile and wearable computers* GR/R98105/01.

References

[1] J. Accot and S. Zhai, "Refining Fitts' law models for bivariate pointing," in Proceedings of ACM CHI, Fort Lauderdale, Florida, 2003, pp. 193-200.

[2] K. Hinkley, J. Pierce, and E. Horvitz, "Sensing Techniques for Mobile Interaction," in Proceedings of ACM UIST, 2000, pp. 91-100.

[3] R. J. Jagacinski and J. M. Flach, *Control Theory for Humans*: Lawrence Erlbaum Associates, 2003.

[4] I. S. MacKenzie, T. Kauppinen, and M. Silfverberg, "Accuracy Measures for Evaluating Computer Pointing Devices," in Proceedings of the ACM CHI, 2001, pp. 9-16.

[5] K. Partridge, S. Chatterjee, V. Sazawal, G. Borriello, and R. Want, "TiltType: Accelerometer-Supported Text Entry for Very Small Devices," in Proceedings of UIST, 2002.

[6] A. Pirhonen, S. A. Brewster, and C. Holguin, "Gestural and Audio Metaphors as a Means of Control for Mobile Devices," in Proceedings of ACM CHI, Minneapolis, 2002, pp. 291-198.

[7] J. Rekimoto, "Tilting Operations for Small Screen Interfaces (Tech Note)," in Proceedings of UIST, 1996.

[8] T. B. Sheridan and W. R. Ferrell, *Man-Machine Systems: Information, Control, and Decision Models of Human Performance*: MIT Press, 1974.

[9] J. Williamson and R. Murray-Smith, "Dynamics and probabilistic text entry," Department of Computing Science, Glasgow University DCS Technical Report TR-2003-147, June 2003.

Mobile Note Taking: Investigating the Efficacy of Mobile Text Entry

Joanna Lumsden[1] and Andrew Gammell[2]

[1] National Research Council of Canada, IIT e-Business, 46 Dineen Drive, Fredericton, New Brunswick, Canada, E3B 9W4
jo.lumsden@nrc-cnrc.gc.ca
[2] Faculty of Computer Science, University of New Brunswick, Fredericton, New Brunswick, Canada, E3B 5A3
andrew.gammell@unb.ca

Abstract. When designing interaction techniques for mobile devices we must ensure users are able to safely navigate through their physical environment while interacting with their mobile device. Non-speech audio has proven effective at improving interaction on mobile devices by allowing users to maintain visual focus on environmental navigation while presenting information to them via their audio channel. The research described here builds on this to create an audio-enhanced single-stroke-based text entry facility that demands as little visual resource as possible. An evaluation of the system demonstrated that users were more aware of their errors when dynamically guided by audio-feedback. The study also highlighted the effect of handwriting style and mobility on text entry; designers of handwriting recognizers and of applications involving mobile note taking can use this fundamental knowledge to further develop their systems to better support the *mobility* of mobile text entry.

1 Introduction

Many experts predicted that the first decade of the 21st century will be the decade of mobile computing: although mobile and wearable computers have been one of the major growth areas in computing in recent years, thus far the promise and hype have surpassed the substance [1]. Why is this? A recent international study of users of mobile handheld devices suggests that there is a predominant perception that quality of service is low and that mobile applications are difficult to use; furthermore, although users give credit to the *potential* of emerging mobile technology, the study highlighted that there is a general feeling that the technology is currently dominating rather than supporting users [2].

Although users are generally forgiving of physical limitations of mobile devices due to technological constraints, they are far less forgiving of the interface to these devices [3]. Despite the obvious disparity between desktop systems and mobile devices in terms of 'traditional' input and output capabilities, the interface designs of most mobile devices are based heavily on the tried-and-tested desktop design paradigm. Desktop user interface design originates from the fact that users are

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 156–167, 2004.

stationary – that is, sitting at a desk – and can devote most (or all) of their attentional resources to the technology with which they are interacting. Users of mobile technology, on the other hand, are typically in motion when they use their devices. This means that they cannot devote all of their attentional resources – especially visual resources – to interacting with their device; such resources must remain with their primary task, often for safety reasons [4]. When designing interaction techniques for mobile devices we must be mindful of the need to ensure that users are able to safely navigate through their physical environment while interacting with their mobile device. It is hard to design visual interfaces that accommodate users' limited attention; that said, much of the interface research on mobile devices tends to focus on visual displays, often presented through head-mounted graphical displays [5] which can be obtrusive, are hard to use in bright daylight, and occupy the user's visual resource [6].

The research presented in this paper is part of an ongoing investigation into how we might improve interaction techniques for mobile devices to better align mobile technologies with human modes of behavior, especially their mobility. Broadly speaking, we aim to enhance the limited existing stylus-based input capabilities to better match the multi-tasking, mobile demands of users as well as to develop new, multimodal interaction techniques for mobile technology and to assess the effectiveness of such techniques. Non-speech audio has proven very effective at improving interaction on mobile devices by allowing users to maintain their visual focus on navigating through their physical environment while presenting information to them via their audio channel [7-10]. The research described here builds on this to create an audio-enhanced single-stroke-based text entry facility that demands as little of users' visual resource as possible, and to assess the effectiveness of such a system.

2 Background

Handwriting recognition systems are one of the primary means of text entry for mobile devices. Handwriting-based interaction is often seen by users as one of the more *natural* text entry techniques, due largely to their prior experience with writing on paper [11]; that said, it is impeded by the fact that users are generally unable to form characters, decipherable to the recognition engine, at rates equal to keyboard tapping [12, 13].

One of the difficulties encountered when using handwriting recognizers is known as the *segmentation problem*: this occurs where the recognizer cannot determine whether a stroke input is intended as part of the previously entered character or as (part of) a new character. Goldberg and Richardson proposed a system called Unistrokes which was designed to avoid this segmentation problem: each character is represented by a distinct, single-stroke gesture which allows characters to be input on top of each other (thereby requiring a greatly reduced writing area) and at the same time – albeit in theory given that their claim was never tested – supporting eyes-free text input [14]. Despite its advantages, the Unistroke system never became widely accepted. Some researchers suggest that this is due to the low correlation between the stroke representation of the various characters and their traditional shape within the Roman alphabet on which they were modeled [15]. The Unistroke *principle*, however, has persisted – most successfully as Palm Inc.'s Graffiti® in which the characters

exhibit a greater degree of correlation with their traditional Roman alphabet representation and for which average accuracy rates (for stationary use) of approximately 96% after only five minutes of use have been reported [13].

While many studies have investigated the usability of different handwriting and single-stroke recognizers and have compared such systems against other alphanumeric input techniques, none of these assessments have addressed the issue of mobility during text entry [11, 13, 14, 16-19]. If mobile devices are to truly support *mobile* activities such as field work, the effect of mobility on the use of these text input techniques needs to be assessed and dealt with accordingly. The research presented in this paper is an initial attempt to establish a corpus of knowledge about the effect of mobility on text entry for mobile technology; it looks at one possible means by which to enhance single-stroke-based text entry to better support *mobile* text input and assesses its effectiveness.

3 Audio-Enhanced Mobile Text Entry

Non-speech audio has proven effective at improving interaction with and presenting information non-visually on mobile devices. For example, Pirhonen *et al.* examined the combined effect of using non-speech audio feedback and gestures to control an MP3 player on a Compaq iPAQ [9]. They designed a small set of metaphorical gestures, corresponding to the control functions of the player, which users could perform – while walking – simply by dragging their finger across the touch screen of the iPAQ. Audio feedback was used to inform users about the completion of their gestures. Pirhonen *et al.* showed that the audio/gestural interface was significantly better than the standard, graphically-based media player on the iPAQ. In particular, the audio feedback upon gesture completion was found to be very important so that users knew what was going on; without it, users' gesture performance was worse than when this feedback was available. Using non-speech audio feedback *during* gesture generation it is possible to improve the accuracy – and *awareness* of accuracy – of gestural input on mobile devices when used while walking [8].

Single-stroke alphabets are gestural in nature and thereby have much in common with the alphanumeric gesture-based work of Brewster *et al.* and Pirhonen *et al.* [8, 9]. Like these gestural systems, single-stroke text entry has the potential to be used eyes-free to input data to a mobile device while walking [14]. Motivated by, and based on, the work of Brewster *et al.* [8] and Goldberg and Richardson [14] together with the fact that Graffiti® has shown potential for general acceptance, we have developed an audio-enhanced single-stroke recognizer which is designed to support text entry when mobile. In his study of user acceptance of handwriting recognition systems, Frankish discovered that although users made conscious changes to their handwriting style in attempts to produce characters that would be more accurately interpreted by the recognizer, such changes produced no significantly noticeable improvement in accuracy [20]. He attributes this to lack of both an effective understanding of the recognition process per se and awareness of what would constitute a more acceptable form. It is hoped that our recognizer will validate the eyes-free capabilities of single-stroke alphabets as mooted by Goldberg and Richardson and – via the audio feedback provided – better inform and support users' attempts to correct their entry of mis-recognized characters.

3.1 Single-Stroke Text Recognition

Our recognizer is based around a conceptual 3x3 grid – see Fig. 1; derived from a publicly available algorithm [21], the co-ordinate pairs that are traversed during a given character entry are condensed into a path comprising the equivalent sequence of grid square ('bin') numbers. The active area of the recognizer's writing pad (i.e. the grid) is 1.3cm x 1.3cm; this size has been shown to effectively support single-stroke text entry for users with motor impairments who, while losing gross motor control, retain some degree of fine motor control [22] – a situation perhaps somewhat akin to writing while walking – and is a size which is commensurate with standard Graffiti® writing pads.

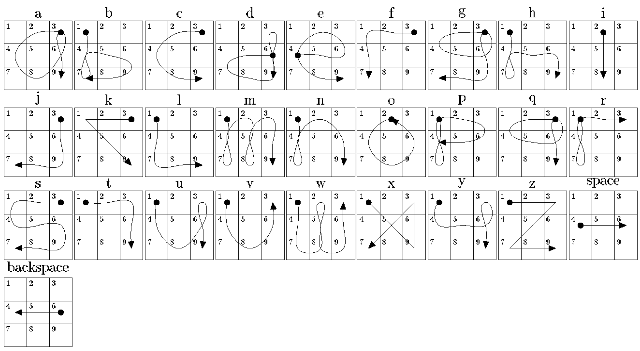


Fig. 1. Single-stroke character set used. Each character is drawn starting at the dot and proceeding to the arrow head along the path shown and each path is unique overall.

For the purpose of our initial investigations, we restricted the character set for use with the recognizer to the 26 lower case letters, space, and backspace as shown in Fig. 1. As can be seen, with the exception of characters that would naturally require more than one stroke to be distinguishable (e.g. 'f', 'k', 't', and 'x'), all characters closely resemble their Roman alphabet representation. For each character, sloppiness space (i.e. error margins defined in terms of acceptable but non-optimal paths) – as defined during pilot testing of the system – was incorporated into the recognition algorithm.

3.2 Sound Design

Sounds were designed to reflect users' interaction with the 3x3 matrix. The sounds were designed to dynamically guide users as they generate textual input as opposed to end-of-entry notification. As part of our investigation, we wished to evaluate the appropriateness of different audio cues; we therefore designed two different soundscapes to enhance the recognizer. In accordance with the findings of Brewster *et al.* [8], we have kept both audio designs as simple as possible to avoid cognitively overloading users. Both designs are based on the C-major chord and all notes are played using the Clarinet timbre (previously proven effective in a gestural context [8]).

1. Bin-Based Audio: This implementation uses a combination of stereo panning and pitch to represent stylus position within the writing pad of the recognizer – see Fig. 2. The note corresponding to the bin row in which the stylus is currently located is played with left panning if in the left-hand column (bins 1, 4 or 7), right panning if in the right-hand column (bins 3, 6 or 9), and equal stereo panning if in the center column (bins 2, 5 or 8). Hence, if a user was to draw a horizontal line from bin 4 to bin 6 (corresponding to the space character in our alphabet), he/she would hear a single tone (C_4) 'move' from left to right. On the basis of this design and the assumption that, in order to be differentiable by the recognizer, no two characters can have the same bin-path, each character also has a distinct audio signature.

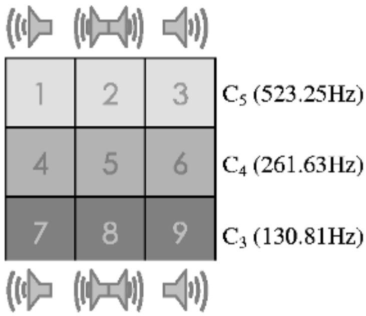


Fig. 2. Bin-based audio design.

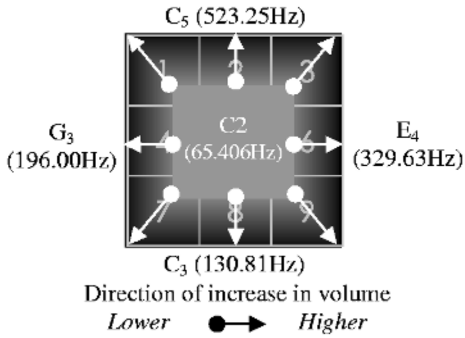


Fig. 3. Boundary-based audio design.

2. Boundary-Based Audio: This implementation moves away from a simple pitch-bin mapping; instead, it attempts to 'reinforce' the virtual boundaries of the recognizer's writing pad using a combination of pitch and relative intensity. Physical boundaries have proven advantageous when used to support single-stroke text entry for users with motor impairment [22]; we wanted to see whether virtual representation of boundaries might have a similar effect on single-stroke text entry for mobile users for whom motor control related to text entry is impeded due to the act of walking itself. As can be seen in Fig. 3, a different pitch is used to represent each of the four sides of the writing pad; as the user draws nearer a boundary of the writing pad, the relative intensity of the tone corresponding to that boundary increases to warn the user of the risk that he/she might slip out of the writing pad. Pitch is used to indicate which of the boundaries the user is approaching; this information can also reinforce to the user his/her direction of movement.

Lumsden *et al.* showed that the *absence* of sound can effectively convey information but only when a sound is anticipated [23]; we needed to enable users to differentiate between the situation where they are in the central zone of the writing pad from the situation where they are outside the writing pad, especially – for eyes-free interaction – when first making contact with the surface. To do this, we introduced a low, unobtrusive tone (C_2) – played whenever the stylus is in the center of the writing pad – as positive reinforcement that surface contact was being maintained as well as allowing the absence of sound to indicate to users that they were *outside* of the writing pad.

4 Experimental Design and Procedure

An experiment was conducted to see whether presenting dynamic audio feedback for textual characters as they are written would, for use in motion, improve users' text entry accuracy and to compare the two sound designs. Additionally, we looked at the degree to which handwriting style and mobility effected the use of the recognizer.

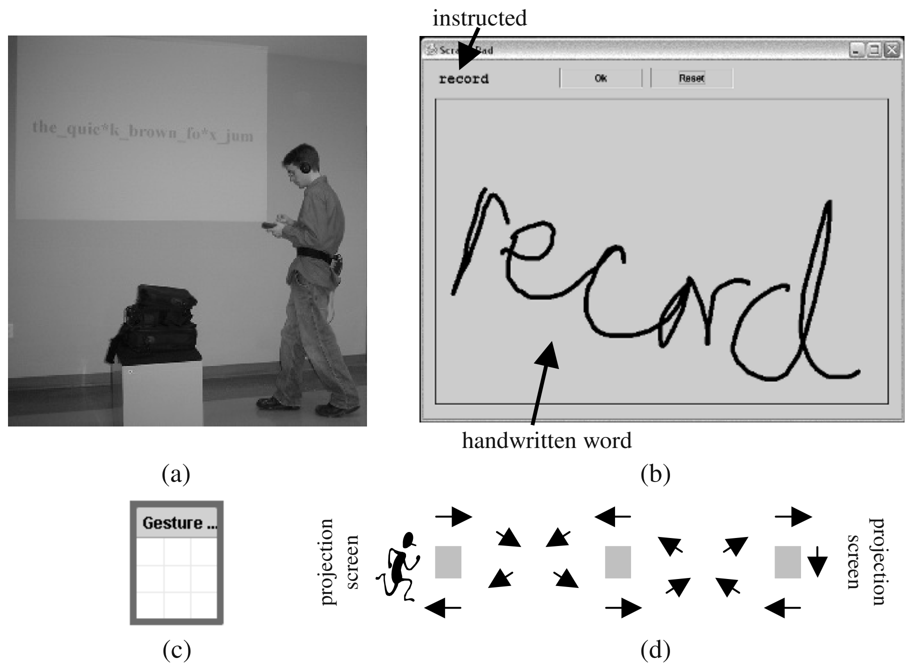


Fig. 4. (a) The wearable computer in use during an experimental session; (b) the handwriting classifier; (c) the text entry pad; and (d) the physical lab set-up.

For the purpose of our experiment, we used a wearable computer (a Xybernaut MA V running Windows XP) which was attached around the participants' waists using a specially designed belt. The single-stroke recognizer (similar in all respects other than feedback across all experimental conditions) ran on the wearable's touch screen which the participants carried in their non-preferred hand; they entered characters using a stylus held in their preferred hand. The recognizer could be positioned (within the display) at the discretion of each user to maximize perceived comfort. Audio feedback was presented to the participants via a pair of lightweight headphones which allowed them to hear the audio output without obscuring real world sounds. Fig. 4(a) shows the equipment in use; the writing pad of the recognizer is shown in Fig. 4(c).

A fully counterbalanced, between-groups design was adopted with each participant performing text entry tasks while walking using the recognizer with no audio feedback and the recognizer with one of the two audio designs. Twenty four people participated (12 per experimental group): 13 females and 11 males ranging in age from 18 to 50 years. Participants were asked to walk 20m laps around obstacles set up in our lab (Fig. 4(d)) – the aim being to test our system while users were mobile in a

fairly realistic environment but maintain sufficient control so that measures could be taken to assess usability. We also asked all participants to perform text entry tasks while seated using the non-audio version of the recognizer; this condition, which was included in the counterbalancing to account for learning affects, allowed us to assess the effect of mobility per se on text entry.

Before embarking on the main component of the experiment, each participant was asked to write, according to their natural handwriting style, a series of 35 English language words. Participants wrote the words, while seated, using the wearable's touch screen and stylus on which we ran a simple drawing surface that captured the 'image' of the participants' handwriting including the number of pen-up and pen-down events per word (see Fig. 4(b)). Using Vuurpijl and Schomakers' categories of handwriting [24], we classified participants' handwriting as *handprint*, *cursive*, or *mixed*.

Brief training was provided prior to each of the three conditions. Participants were given printed training material which outlined how to enter each of the 28 characters used – this was identical for all three conditions; condition specific explanation of the audio feedback design was included for the two audio conditions. Participants were then given 5 minutes of practical use of the recognizer during which to familiarize themselves with the current version; during the latter 3 minutes, participants were asked to enter text according to the requirements of the actual experimental session.

During each condition, participants were asked to enter ten 4-word English language phrases (selected, as far as possible, from the set proposed by MacKenzie and Soukoreff [25]). Each phrase was projected onto the wall at one or other end of the circuit (see Fig. 4(d)) at random; participants were asked to locate the projected phrase and enter it using the writing pad. The results of participants' text entry – that is, the recognized characters – were projected onto the opposite wall to the original phrase (see Fig. 4(a)); no visual representation of their input was provided on the touch screen. Input that was undecipherable to the recognizer was represented with an '*' in the projected output sequence. When participants completed a phrase, they hit a 'Submit' button on the touch screen and the next phrase was projected; for the two mobile conditions, participants were asked to enter one phrase per physical lap of the circuit. We adopted this set-up to force participants to look up from the touch screen as they entered text – as users would have to do in a less stable physical environment – as well as to introduce a level of distraction (projected phrases and output representation were not always in the same place and participants were not always directly facing what they needed to see) in an attempt to reflect real world situations as much as possible in a lab setting. Three different phrase sets were used during the course of each experiment. The order of use of the phrase sets remained constant while the condition order was counter-balanced; this was done to eliminate any potential bias that may have arisen due to some phrases being perceived as 'easier' than others.

During the experiment, a full range of measures, including accuracy rates and subjective workload (using the NASA TLA [26] scales), was taken to assess the usability of the audio designs tested and to investigate the effect of handwriting style and mobility on the use of our system. It is important to consider workload in a mobile context: users must split their attentional resources between their physical environment and tasks with which they are engaged (both technology-based and otherwise) and so any interface that can reduce workload is more likely to succeed in a real mobile setting.

To assess the difference in the degree to which the various versions of our recognizer affected users' walking speed, we also recorded percentage preferred walking speed (PPWS) [27]; the greater the impact on users' walking speed, the less effective the audio designs were at supporting eyes-free text entry. Pirhonen *et al.* found this to be a sensitive measure of the usability of a mobile device – in their case, a mobile MP3 player [9]. Prior to the start of each experiment, participants walked 5 laps of the room wearing all the equipment; their lap times were recorded and averaged so that we could calculate their standard PWS when carrying, but not interacting with, the technology.

The main hypotheses were that mobility would have a significantly detrimental effect on text input accuracy using a single-stroke alphabet and that, when mobile, users would input text more accurately under the audio conditions than non-audio condition. It was also hypothesized that, as a result of increased cognitive load, the audio-conditions would have a greater detrimental effect on participants' PWS than the non-audio condition when mobile. Since both audio conditions were previously untried, we made no hypothesis as to which would return better results. Our final hypothesis was that users whose natural handwriting style fell into the *handprint* category would outperform those whose handwriting was classified as *cursive* or *mixed*; this was on the basis that *handprint* (i.e., writing that averages one pen-stroke per character) appears to have greater affinity with the requirements for text entry using a single-stroke ('print' style) alphabet than the other categories of handwriting.

5 Results and Discussion

A two factor ANOVA showed that experimental condition significantly affected participants' subjective assessment of overall workload ($F_{2,54}$=4.20, p=0.020). Tukey HSD tests showed that participants experienced significantly less workload when seated than when mobile under both the audio conditions (p=0.032) and non-audio condition (p=0.04). There was no significant difference observed between the audio and non-audio mobile conditions. Of the six dimensions of workload, only two were shown to be significantly different across the experimental conditions. A two factor ANOVA confirmed that Physical Demand was significantly greater when mobile than seated ($F_{2,54}$=6.44, p=0.003) with both the audio mobile and non-audio mobile conditions imposing significantly more physical demands than the seated condition (p=0.01 and p=0.001 respectively). There was no significant difference in terms of Physical Demand observed between the audio and non-audio mobile conditions. Hence, rather unsurprisingly, mobility has been shown to increase the experience of workload for text entry. A two factor ANOVA showed that experimental condition had a significant impact on participants' self assessment of Performance ($F_{2,54}$=3.80, p=0.029). Participants' rated their performance significantly lower when mobile using the audio versions of the recognizer than when seated using the silent version (p=0.0235); there were, however, no significant differences between the audio and non-audio mobile conditions nor between the non-audio mobile and seated conditions. At the level of conjecture, this may be due to the fact that participants were more aware of their errors when given audio feedback (see below) and so better placed to assess their performance (accuracy averaged 83% for seated use and 78% for mobile use).

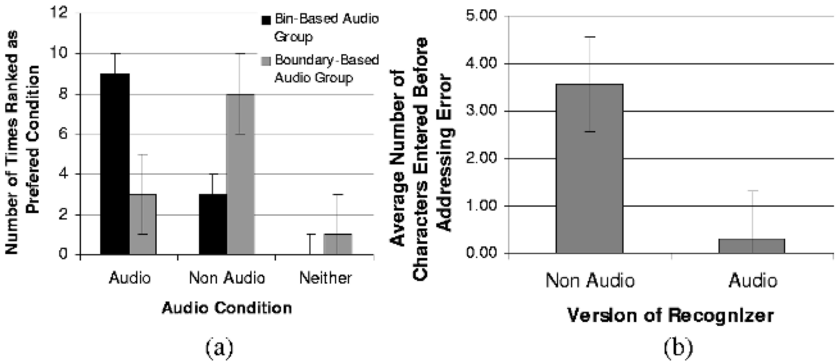


Fig. 5. Recognizer version: (a) stated preference according to experimental group; (b) awareness of error.

A two factor ANOVA showed that, for mobile use, the combination of experimental group and condition significantly affected participants' stated preference of recognizer ($F_{1,44}$=11.78, p=0.001). Participants in the experimental group using the bin-based audio version of the recognizer significantly preferred using the recognizer with audio feedback than without ($p<0.05$). The difference in preference within the group using the boundary-based audio version of the recognizer was not significant. Participants' allocation of preference is shown in Fig. 5(a). This observation is particularly interesting in light of the accuracy results (see below for further discussion); across all conditions, the accuracy results for the group using the bin-based audio were significantly higher than for the other group and it is this same group that preferred the audio version of the recognizer. While we would hope there to be a link between these findings – i.e., that participants in the first group preferred the bin-based audio version of the recognizer *because* they subjectively felt it improved their accuracy – further evaluation would be required to confirm this. Handwriting style was not shown to significantly influence preference.

Several factors were shown to significantly affect the accuracy of text entry which we measured using Soukeroff and MacKenzies' Unified Error Metric [28]. A two factor ANOVA showed that handwriting style had a significant effect on participants' accuracy ($F_{2,54}$=3.30, p=0.044) with *cursive* hand writers making significantly fewer errors than participants who *handprint* (p=0.035). No significant differences were otherwise observed in terms of handwriting style. This observation was surprising given our initial hypothesis; further investigation will be required to assess *why* the *cursive* style corresponds to more accurate entry.

We observed group allocation to significantly affect the accuracy of participants' text entry ($F_{1,54}$=6.34, p=0.015) with the participants in the group using the bin-based audio making significantly fewer errors than participants in the other group (p=0.014). The specific audio design was also observed to significantly affect participants' text entry ($F_{1,22}$=4.84, p=0.039); using the bin-based audio design, participants made significantly fewer errors than participants using the boundary-based audio design ($p<0.05$). We cannot, from the results obtained, determine cause and effect – i.e. was it the audio design that encouraged participants in the bin-based audio group to be more accurate per se or were the participants in this group, despite being randomly selected and assigned to the group, predisposed to be more accurate?

Given the similarity between the bin-based audio design and the work of Brewster *et al.*[8] we would like to attribute the superior accuracy to the audio rather than the people, but will have to conduct further evaluations to determine whether this is a valid assumption. We can, however, conclude that the bin-based audio was the more effective audio design. We found no significant difference between accuracy of text entry across the audio mobile and non-audio mobile conditions.

To assess participants' level of *awareness* of their text entry when mobile, we measured the average number of characters participants entered following an error before realizing and addressing (by deleting the erroneous partial and/or complete character(s) entered) their mistake (see Fig. 5(b)). We found the availability of audio feedback to significantly affect awareness ($F_{1,22}$=6.65, p=0.015); when mobile, given audio feedback, participants entered significantly fewer characters before realizing and addressing mistakes than when the audio feedback was absent ($p<0.05$). This suggests that the audio feedback *increased* error awareness *during* erroneous entry; without audio feedback, participants had to rely on visual identification of errors which was less efficient/effective given competing demands on their visual resource.

Contrary to our hypothesis, the audio versions of the recognizer were not found to significantly reduce participants' walking speeds compared to the non-audio version. Although all mobile conditions had a noticeably detrimental impact on participants' walking speeds when performing the text entry tasks (speeds ranged from 28% to 31% of PWS), tests showed audio condition to have no significant effect on PPWS. Similarly, handwriting style had no significant effect on PPWS.

6 Conclusions

This paper has shown that handwriting-based interaction techniques that combine sound and gesture have significant potential to support mobile note taking. Audio feedback has been shown to significantly improve users' awareness of errors made during mobile text entry. Of the two soundscapes evaluated, the bin-based audio design was preferred to the boundary-based audio design and supported more accurate text entry when mobile. This suggests that the simpler the design, and the more direct and immediate the mapping between feedback and user gesture, the better (to avoid overloading users' auditory and cognitive capacity). This improvement in awareness was not, when compared to the effect of the non-audio version of the recognizer, at the expense of walking speed nor at the detriment of workload.

Handwriting style was shown to significantly affect users' text entry accuracy which implies that there is potential benefit in investigating how to better support handwriting recognition, in particular in motion, based on tailorability to style.

Since users only achieved average accuracy rates when performing mobile text entry that were 20% *below* the recognized acceptance rate for stationary use of handwriting recognition systems, there remains considerable scope for further investigation and improvement in this regard.

We have, however, shown that it is possible to support mobile note taking using techniques that allow, to a greater degree than would otherwise be feasible, for eyes-free text entry. Designers of handwriting recognition systems and of applications to

support activities involving mobile note taking now have a basis of knowledge upon which to further develop their systems to better support the *mobility* of mobile text entry.

Acknowledgements. This work has been funded in part by NSERC Discovery Grant 262111-03.

References

1. Urbaczewski, A., Valacich, J.S., and Jessup, L.M., *Mobile Commerce: Opportunities and Challenges.* Communications of the ACM, **46**(12), (2003), p. 30 - 32.
2. Jarvenpaa, S.L., Lang, K.R., Takeda, Y., and Tuunainen, V.K., *Mobile Commerce at Crossroads.* Communications of the ACM, **46**(12), (2003), p. 41 - 44.
3. Sarker, S. and Wells, J.D., *Understanding Mobile Handheld Device Use and Adoption.* Communications of the ACM, **46**(12), (2003), p. 35 - 40.
4. Brewster, S.A., *Overcoming the Lack of Screen Space on Mobile Computers.* Personal and Ubiquitous Computing, **6**(3), (2002), p. 188 - 205.
5. Barfield, W. and Caudell, T., *Fundamentals of Wearable Computers and Augmented Reality*, Mahwah, New Jersey: Lawrence Erlbaum Associates, (2001).
6. Geelhoed, E., Falahee, M., and Latham, K., *Safety and Comfort of Eyeglass Displays*, in *Handheld and Ubiquitous Computing*, P. Thomas and H.W. Gelersen, Editors, Springer: Berlin, (2000), p. 236 - 247.
7. Holland, S. and Morse, D.R. *Audio GPS: Spatial Audio Navigation With a Minimal Attention Interface*, in *Mobile HCI 2001: Third International Workshop on Human-Computer Interaction with Mobile Devices*, Lille, France, (2001), p. 253 - 259.
8. Brewster, S.A., Lumsden, J., Bell, M., Hall, M., and Tasker, S. *Multimodal 'Eyes-Free' Interaction Techniques for Mobile Devices*, in *Human Factors in Computing Systems - CHI'2003*, Ft Lauderdale, USA, (2003), p. 473 - 480.
9. Pirhonen, P., Brewster, S.A., and Holguin, C. *Gestural and Audio Metaphors as a Means of Control in Mobile Devices*, in *ACM-CHI'2002*, Minneapolis, MN: ACM Press Addison-Wesley, (2002), p. 291 - 298.
10. Sawhney, N. and Schmandt, C., *Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments.* ACM Transactions on Computer-Human Interaction, **7**(3), (2000), p. 353 - 383.
11. MacKenzie, I.S., Nonnecke, B., Riddersma, S., McQueen, C., and Meltz, M., *Alphanumeric Entry on Pen-Based Computers.* International Journal of Human-Computer Studies, (41), (1994), p. 775 - 792.
12. Zhai, S. and Kristensson, P. *Shorthand Writing on Stylus Keyboard*, in *Human Factors in Computing Systems - CHI'2003*, Ft. Lauderdale, USA: ACM Press, (2003), p. 97 - 104.
13. MacKenzie, I.S. and Zhang, S. *The Immediate Usability of Graffiti*, in *Graphics Interface '97*, Toronto: Canadian Information Processing Society, (1997), p. 129 - 137.
14. Goldberg, D. and Richardson, C. *Touch-Typing With a Stylus*, in *Conference on Human Factors in Computing Systems - INTERCHI '93*, Amsterdam, The Netherlands: ACM Press, (1993), p. 80 - 87.
15. MacKenzie, I.S. and Soukoreff, R.W., *Text-Entry for Mobile Computing: Models and Methods, Theory and Practice.* Human Computer Interaction, **17**, (2002), p. 147 - 198.
16. MacKenzie, I.S., Nonnecke, R.B., McQueen, J.C., Riddersma, S., and Meltz, M. *A Comparison of Three Methods of Character Entry on Pen-based Computers*, in *Proceedings of the Human Factors and Ergonomics Society 38th Annual General Meeting*, Santa Monica, CA: Human Factors Society, (1994), p. 330 - 334.

17. LaLomia, M.J. *User Acceptance of Handwritten Recognition Accuracy*, in *Human Factors in Computing Systems - CHI'94*, Boston, MA, USA: ACM Press, (1994), p. 107.
18. Frankish, C., Hull, R., and Morgan, P. *Recognition Accuracy and User Acceptance of Pen Interfaces*, in *Human Factors in Computing Systems - CHI'95*, Denver, USA, (1995), p. 503 - 510.
19. MacKenzie, I.S. and Chang, L., *A Performance Comparison of Two Handwriting Recognizers.* Interacting With Computers, **11**, (1999), p. 283 - 297.
20. Frankish, C., *Pen-Based Computing*, in *Interface Technology: The Leading Edge*, J.M. Noyes and M. Cook, Editors, Research Studies Press Ltd: Baldock, England, (1999), p. 59 - 72.
21. Willey, M., *LibStroke - A Stroke Translation Library*, ETLA Technical Services, http://www.etla.net/~willey/projects/libstroke/, (2002).
22. Wobbrock, J.O. *The Benefits of Physical Edges in Gesture-Making: Empirical Support for and Edge-Based Unistroke Alphabet*, in *Human Factors in Computing Systems - CHI'2003*, Ft. Lauderdale, USA, (2003), p. 942 - 943.
23. Lumsden, J., Brewster, S.A., Crease, M., and Gray, P.D. *Guidelines for Audio-Enhancement of Graphical User Interface Widgets*, in *Proceedings of BCS HCI'2002*, London, (2002).
24. Vuurpijl, L. and Schomaker, K. *Coarse Writing-Style Clustering Based on Simple Stroke-Related Features*, in *5th International Workshop on Frontiers in Handwriting Recognition*: IAPR, (1996), p. 29 - 34.
25. MacKenzie, I.S. and Soukoreff, R.W. *Phrase Sets for Evaluating Text Entry Techniques*, in *Human Factors in Computing Systems - CHI'2003*, Ft. Lauderdale, USA, (2003), p. 754 - 755.
26. Hart, S.G. and Wickens, C., *Workload assessment and prediction*, in *MANPRINT: an approach to systems integration*, H.R. Booher, Editor, Van Nostrand Reinhold: New York, (1990), p. 257 - 296.
27. Petrie, H., Furner, S., and Strothotte, T. *Design Lifecycles and Wearable Computers for Users with Disabilities*, in *The First Workshop on Human-Computer Interaction with Mobile Devices*, Glasgow, UK, (1998).
28. Soukeroff, R.W. and MacKenzie, I.S. *Metrics for Text Entry Research: An Evaluation of MSD and KSPC, and a New Unified Error Metric*, in *Conference on Human Factors in Computing Systems (CHI'2003)*, Ft Lauderdale: ACM Press, (2003), p. 113 - 120.

The Personal Audio Loop: Designing a Ubiquitous Audio-Based Memory Aid

Gillian R. Hayes, Shwetak N. Patel, Khai N. Truong, Giovanni Iachello, Julie A. Kientz, Rob Farmer, and Gregory D. Abowd

College of Computing & GVU Center
Georgia Institute of Technology
Atlanta, GA 30332-0280, USA
{gillian, shwetak, khai, giac, julie, rfarmer, abowd}@cc.gatech.edu

Abstract. Sound is an important medium in our lives, but its ephemeral nature can be problematic when people cannot recall something they heard in the past. Motivated by everyday conversational breakdowns, we present the design of a continuous, near-term audio buffering application: the Personal Audio Loop (PAL). PAL was designed as a truly ubiquitous service to recover audio content from a person's recent past. Initial brainstorming and prototyping for PAL revealed major aspects of the design space that require further investigation, including potential usefulness in everyday life, the level of ubiquity required, the usability features for any instantiation of the service, and the social and legal considerations for potential deployment. We present a design of PAL, informed by a controlled laboratory study, diary study, and examination of pertinent legislation. We conclude with an analysis of the results and some initial observations of the deployment of a prototype developed for a Motorola i730 handset.

1 Introduction

Everyday conversations fill our lives, and we are all very familiar with the kinds of breakdowns suggested by these simple scenarios:

– You are in a conversation with a friend, and one of you is interrupted. When the conversation resumes, neither of you can remember what you were talking about.
– You are at a social event, and you are introduced to someone new. Minutes later, you have forgotten the person's name.

We have a particular interest in automated capture of live experiences for later access, and we are naturally drawn to these scenarios, because they demonstrate the use of audio capture with near-term access. Over the past three years, we have experimented with different technical approaches, and have found that a mixed technological and human-centered approach is necessary to produce a near-term (*i.e.*, less than one day) audio service that would be likely to survive a real deployment. Such a design must answer questions of human significance pertaining to the following issues:

– *Usefulness:* Though motivated by observations from everyday life, how often and in what situations do people actually need a near-term audio memory aid?
– *Ubiquity:* What parameters of such a service would make it available everywhere and every time someone needed it?

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 168–179, 2004.

– *Usability:* How should the service deliver functionality to maximize its benefit and minimize its distraction?
– *Social and legal considerations:* What social and legal concerns might prevent the successful deployment of an audio recording application for everyday life?

An automatic audio-based memory aid is arguably outside of the realm of a typical person's experience. Therefore, potential users should be able to interact with a working prototype to have a sense of the capabilities, necessitating the answering of engineering questions as well, including important architectural considerations.

From a technical perspective, there are several options for designing an audio-based memory aid to provide the capability motivated by the above examples. Although all designs reflect the same basic notion of replaying a buffer of recently recorded audio, early prototypes varied in terms of distribution of recording and playback capabilities. A fully distributed system assumes an instrumented environment, with microphones, speakers and interface controls placed to maximize opportunities for recording and playback wherever and whenever needed. A fully localized solution provides recording and playback in an all-in-one package carried wherever needed. A hybrid solution might allocate the recording in the environment and accomplish playback through a handheld device that receives streamed audio from a central repository.

In this paper, we present a design study of the Personal Audio Loop (PAL), a solution for a deployed near-term audio reminding service that addresses both the technical concerns of an interesting capture and access application while also answering questions from the four categories described above. The process involved a series of formative studies that led to the design of a self-contained service integrated into a commercial mobile phone handset. Although the decision to build a local solution for PAL came fairly early, it results naturally from an exploration of the usefulness, ubiquity and socio-legal concerns for this problem, and it is justified by our findings.

In the next section, we provide a brief background of technology and of relevant social and legal work in this area. In Section 3, we describe the initial implementation of PAL on a commercial mobile phone handset and outline the various empirical and diary studies that formed the basis for our formative studies. In Section 4, we give preliminary results from an initial deployment study and in Section 5 we summarize the critical design features of PAL. Finally, in Section 6 we summarize the contributions of this work and outline future work.

2 Background and Related Work

Near-term capture and access applications that provide audio reminder services have been previously explored in the office as well as for telephone conversations. Xcapture, originally built to provide a "digital tape loop" of a single office, could also provide short-term auditory memory of telephone conversations (5 to 15 minutes long) [9]. Although the system was designed for use in a setting where social protocol allows recording, the authors recognized the privacy issues of subsequent use of archived recordings, and suggested that social expectations change with use. In MERL's real-time audio buffering technique, captured audio persists for the duration of that phone conversation [4]. During the course of the conversation, a user may tap

the phone against the ear to move backwards in the audio and to replay any portion of the discussion. This system does not store conversations and could arguably pass legislative tests and be socially acceptable. Video has been employed for reminder services as well: the Déjà vu Display (previously known as the Cook's Collage) explores the use of collage displays to show recent activities in the kitchen [13]. In the case of a memory lapse or interruption, the user can rely on annotated snapshots of key steps to remind her of the last few things she did. Although the Déjà vu Display was designed for private space (*i.e.* the home), much attention was given to specific privacy-friendly affordances, such as the camera angle, the richness of captured data and avoiding sound recording. Such affordances are determinant in the equilibrium of privacy and will be extensively discussed below.

Legal cases over the past two decades have exposed the contrasting requirements and balances of privacy and utility for recording applications. We draw from the experience in the fields of surveillance in public spaces and of the privacy of private communications.[1] Among other sources we considered European Directive 95/46/EC, [6] together with opinions and rulings by various EU Data Protection Authorities (DPAs) [5, 2] and several US Supreme Court[2] rulings—the most relevant being the *Katz v. United States* [10] case, which extended the right of privacy to what the individual seeks to protect from the public and the *Kyllo v. United States* [11] case, which indicated that the subjects of surveillance are granted a sufficient expectation of privacy if the surveillance technology employed is not in common use.[3]

Despite the ongoing debate stressing the differences between the United States and Europe regarding privacy, legislation regulating the recording of communications by electronic means is remarkably similar. The main items are the US Electronic Communications Privacy Act (ECPA) of 1986 [14] and European Directives 2002/58/EC [7] and 95/46/EC. ECPA regulates wiretap and surveillance and applies to any electronic recording device and conversations ("oral communication") between two persons "exhibiting an expectation that such communication is not subject to interception," even if the conversations were not transmitted through a telecommunications network. European Directive 2002/58/EC covers only personal conversations transmitted over public telecommunication networks. However, Directive 95/46/EC applies to any personally identifiable information, which includes recorded voice conversations, according to multiple opinions by European national data protection authorities. Although the Directive was originally meant to regulate the management of personal data collected by organizations in large textual databases, recent opinions expressed by DPAs have addressed cases of more limited balancing of individuals' rights. As is detailed below, Directive 95/46/EC requires a proportionality assessment between potential harm and benefits; however, the personal character of the application might exempt users from many provisions, including informed consent.

---

1 Most industrialized nations have pertinent legislation; we limit our inquiry to the US Federal legislation and European Union directives. Note that these laws are not directly comparable: US legislation gives states less discretion than EU law gives to member states.

2 The United States do not have DPAs specifically appointed to examine privacy issues.

3 Further information on the details of these and other US Supreme Court decisions can be found at http://www.findlaw.com/casecode/supreme.html.

3 Formative Studies of PAL

Based on early interviews and our intuition, we determined that the platform for PAL would need to be mobile, powerful both in processing and development environment, include buttons, and an external or attachable microphone. The mobility, ubiquity and performance of mobile phones make them an appealing platform for this application, but only certain phones support the required capabilities. Our choice, the Motorola iDEN i730 (Fig. 1) is a clamshell phone featuring a J2ME programming environment conforming to the MIDP 2.0 and Mobile Media APIs. The i730 microphone is capable of recording voices in a small room with the phone open or closed in a shirt pocket or attached to a belt, with higher quality than most PDAs. The two formative studies reported were designed to answer questions of the feasibility of using a mobile phone as the interface to an audio-based memory aid and to characterize the frequency and situations of use in everyday life.

3.1 Laboratory Study: Developing a Usable Phone Interface

In its normal operating mode, our implementation of PAL continuously records audio from the user's environment. Audio older than the buffer length (in our initial prototype, 15 minutes) is automatically deleted. Recording automatically halts when the user answers or makes a call. Five buttons are available on the outside to accommodate interactions while the phone is closed (Fig. 1). PAL provides simple audio navigation features (*e.g.* rewind), informed both by previous research on skimming [1] as well as by commercial video recording services like Tivo™. PAL includes a simple timeline visualization on the exterior LCD of the handset indicating application status (recording, playback and direction of navigation) as well as the playback position in the audio buffer relative to the current time (the right edge of the timeline).

We designed a laboratory study using an early prototype to test the usability of the interface from a quantitative performance perspective and a qualitative impression.

Fig. 1. The Motorola i730 handset used for PAL. Three buttons control navigation and record/playback mode. A timeline indicates mode and relative place in the buffer.

Method. The laboratory study included 18 participants (students and faculty from our institution specializing in HCI research, 5 female, 13 male, ages ranging from 18 to 50). Participants with an HCI background were explicitly chosen with the intent of examining heuristics such as the mapping of buttons to functionality and the quality of the visualization. Participants' experience with mobile phones ranged from seven years of consistent use to no experience at all (7 participants). We demonstrated the prototype, encouraging participants to examine the device and ask questions until they expressed comfort with its functions.

PAL's intended use involves the replay of audio for which the user was present initially. The controlled study, designed to mimic this scenario, included a scripted dialog of five minutes. In this script, the participants asked researchers predetermined questions, and researchers replied with the same answer for every participant. The script purposely involved a large amount of detail to increase the likelihood that participants could not recall the answers to all questions from memory. After completing the dialog, the researchers who had been participating in the dialog removed the script and asked the participants a series of questions about the information they had just been provided. Although it was noted whether the participants remembered the information without use of PAL, every participant was asked to find and play every answer. Participants were encouraged to "think aloud" as they used the prototype, and the researchers timed how long it took an individual to find the answer, theorizing that this first time use while discussing their actions would be a worst case timing for most users. Participants answered seven questions, the first two being practice questions not used for computing timing results. An exit survey and semi-structured interview provided a qualitative evaluation of the interface and of their need for this kind of service.

Results. After a short demonstration, all participants were able to navigate the audio well enough to answer our questions. They commented that the device was easy to use with one hand ($\mu = 6.95$, $\sigma = 0.2$, 7 being the highest), and small enough to carry at all times ($\mu = 5.42$, $\sigma = 2.0$ out of 7). They could clearly understand the audio even in its highly compressed form ($\mu = 6.5$, $\sigma = 0.9$, with 7 being "strongly agree").

With an audio buffer of 15 minutes, participants required an average of 34.8 seconds ($\sigma = 22.58$) to find responses for questions that were known to be in the in the recorded audio while talking aloud about their actions. Participants reported the visualization was somewhat helpful in accomplishing the tasks, but not overwhelmingly so ($\mu = 5.21$, $\sigma = 1.4$, with 7 being "very helpful"). Thirteen out of our eighteen participants used PAL without the visualization, preferring an eyes-free interaction.

Although inquiring about privacy was not a goal of this study, ten of our participants raised spontaneous concerns regarding the social acceptability of a continuously recording system. The most common sentiment expressed indicated that participants were less concerned about recording their own voice than their conversation partners'.

3.2 Diary Study: Determining the Usefulness of PAL

The laboratory study showed the feasibility and usability of PAL on a mobile phone, but it did not inform us about the overall usefulness in everyday life. We undertook a

diary study to explore the extent to which a near-term audio reminder service was needed, looking for frequency and characteristics of potential use. Diary studies balance the ecological validity of gathering such data *in situ* against interruption of everyday activity flow caused by recording personal observations, particularly in mobile settings [3]. We asked for specific information relating to social context including privacy concerns in the diary entries and during the follow-up interviews.

Method. Twelve experienced mobile phone users (5 female, 7 male, ranging in age from 22 to 60 years) participated in the study. Participants' occupations spanned a spectrum of domains, including a psychologist, finance manager, realtor, car dealer, consultant, professor, and full-time homemaker. We demonstrated a fully working version of PAL to participants. We then asked them to carry small pocket-sized diary and record an entry in it for each incident during the following week when they would have needed or liked to use the PAL service. Each page of the diary contained a simple form to complete for the potential instance of use, streamlined after an initial trial period. Each form in the diary included space for describing the content of the audio to retrieve, when and where the incident occurred and whether any persons unrelated to the conversation were nearby. Participants also estimated how far in the past the salient audio content was and rated how important it was to retrieve that information. Fig. 2 shows an example of an incident survey.

At the end of each week, we collected the diaries from participants and conducted semi-structured interviews to examine in detail up to six diary entries per participant per week, including privacy-related questions such as the kind of information being sought, the distance of unrelated third parties from the participant and their assessment of the social appropriateness of using the device in the specific context. We then gave each participant who chose to continue for another week a new empty diary to again record incidents. At the end of the study, we conducted semi-structured interviews with all participants. The weekly and summary interviews allowed us to clarify misunderstandings in the entries as well as to probe particular issues, such as privacy concerns, that were not easily gathered in the chosen diary form factor.

Day of Week	S	M	T	W	T	F	S
Time of Day	12 am - 6 am	6 am - 9 am	9 am - 11 am	11 am - 1 pm	1 pm - 6 pm	6 pm - 9 pm	9 pm - 12 pm

Location: Kroger — Unrelated People Present? N

1) What was the audio you needed? 2) Why?

While shopping I ran into an old friend. We talked for a while and he gave me his phone number. I could not remember it later

How long ago was the audio of interest?	< 10 min	10 min - 1 hour	1 hour - 1 day	> 1 day

How problematic was it to forget this?	1 Neutral	2	3 An annoyance	4	5 Problematic

Fig. 2. Sample diary entry.

Results. Twelve people participated in the first week, eleven of them continued for the second, and eight in the third, for a total of 31 participant weeks and 109 incident reports. Participants reported an average of 3.5 ($\sigma = 2.7$) incidents per week, of which 32% referred to audio from "less than 10 minutes ago", 26% from "10 minutes up to an hour", while only 6% were from over a day prior.

Of the incidents reported, 25% occurred in public, 44% in semi-public spaces (defined as schools, workplaces, *etc.*) and the remaining 31% in private space (predominantly car and

home). In 44% of the incidents, participants indicated that people unrelated to the audio they wished to retrieve (*e.g.*, other customers in a restaurant) had been present during the time they would have liked to record. We collected follow-up information for 83 incidents during the weekly interviews. Participants asserted that they would not have felt rude towards their communication partner using PAL in 52 of these. During the second and third weeks participants were questioned about their reactions had their partners objected to their use of the application. Participants stated that such an objection would be "not likely" in 24 of 26 incidents queried and indicated that they would not have complied with the objection, had there been one, in 19 of the 26 incidents queried. Only in 4 occasions participants asserted that unrelated bystanders could have been concerned had they known that they were using PAL. When asked how far away they would like PAL to record, 67% chose within a small room (10 feet), 22% preferred smaller areas (own voice or arm-length distance), and only one individual requested a large radius, reporting that he "is just nosy".

During interviews, participants reported on how long they would be willing to search for content rated at various levels of problematic. If they were "neutral" (scoring a one on a five point scale) about the content, they reported being willing to spend an average of 336 seconds (σ = 172) to search, whereas if the audio content was of vital importance (scoring a five), they reported being willing to spend a minimum of 15 minutes with three users responding "however long it takes" to retrieve it.

4 Preliminary Results from Deployment

We deployed a working version of the application to four of the diary study participants for seven weeks. Although we do not report on their use, four members of our research team have also been using PAL for over two months. During the first four days of the deployment, we asked participants to carry a diary to note their uses of the device. These participants used the device on average 2.5 times per week (σ = 1.9, pro-rated given the short term of the study). Although this average is lower than what was indicated by the diary study, participants also reported on average 1.5 incidents that they thought about using the device and chose not to (σ = 0.6). In one case, the user's conversation partner recovered the information before the user was able to try with PAL. In all other cases, the reason not to use PAL was reported as forgetting it was available. Informal interviews with the users since this initial probe indicate that ordinary use subsequently remained fairly consistent with the rate observed in the first four days, and that the frequency of use for exploring the application or showing it to others has decreased substantially. Overall, satisfaction as reported through qualitative interviews has been high. All four users requested to continue using the devices after the first four days and reported that they believed they would use them more over time. Each user changed the buffer length (ranging from ten minutes to sixty), the initial jump backward (ranging from 15 seconds to 60), or both. Users expressed that configuring the application was important and one user even indicated that he changes the buffer length depending on the situation he is about to encounter.

By deploying the devices to even a small number of users, we expected to be able to observe uses both expected and emergent and gain greater understanding about the

dependency users might have developed on the service. In the initial four day probe, the most frequent reported situation for use was to remember forgotten details (60%). Other unexpected situations have also been reported in the following weeks. Specifically, users have been employing PAL as an instructional aid: recording conversations with customers and then replaying them for an employee in training. One user has also been using it as a medical journal to record data about symptoms requested by her doctor by speaking them aloud when she can not write them down at that moment. Users have begun to expect the service to be available, reporting that they choose not to write information down when it is already spoken aloud.

Social contract issues recurred more often than the results of the diary study had revealed: users expressed that conversation partners aware of the device sometimes responded negatively initially, but relaxed after the application and its buffering and discarding functions were explained. Interestingly, all four users reported informing new conversation partners about PAL less frequently as time went by. After several weeks, users have almost stopped alerting conversation partners altogether. As frequently as users reported negative social repercussions from PAL, they also reported positive cooperative uses of the device. For example, one user's wife consistently uses PAL on his device by walking near to him and speaking when she needs to remember something. We are exploring in depth the changing behaviors of individuals around the owner as the study continues.

5 Critical Features for Use

Informed by the exploration of privacy regulations and by findings from the laboratory and diary studies, we uncovered the critical features of PAL outlined previously.

Making PAL useful. Given the rates of 2.5 and 3 incidents per week as reported by the deployment and the diary study, the need for PAL is justified. Analysis of the stated purpose for recovering the audio provided additional information, synthesized in Table 1. From the legal perspective, the frequency and unpredictability of use of this application could support a positive argument for the proportionality test (as used in [5]) with regards to the issue of continuous automatic recording.

Table 1. Purpose for recovering audio (total 109 entries in diary study)

Purpose category	Occurrences
Forgotten previous details (*e.g.*, making a list, retrieving details)	36 (33%)
Replaying for conversation partner (replaying for person who either spoke the audio originally or was present to hear it)	20 (18%)
Interrupted (external activity took focus away from important audio)	18 (17%)
Explicit tape recorder behavior (participant was aware prior to the incident that she wanted to record it)	13 (12%)
Distracted (another concurrent activity took attention)	13 (12%)
Relaying information from one partner to another (replaying for person not present when original audio was recorded)	9 (8%)

Information minimization requires collecting the minimum amount of personal information needed by the application. Given that 58% of the diary incidents referred to content within one hour, a buffer up to 60 minutes should suffice, with a 15 minute default. EU and US law diverge in this regard, as ECPA does not make any distinction based on stored information retention time. A more conservative way of looking at this issue would be that of understanding the duration of the "social contract", implicit among parties engaged in a conversation, to determine how long a recording can be maintained after the end of such conversation. This measure relates to the relation between distance and place (in the sense of [8]): how long does it take to move between places with incompatible social contracts? Because PAL could be abused when crossing place boundaries, the recording should be limited to minimize such risks. While valid from a phenomenological standpoint, we decided to postpone this assessment, given the unsolved issue of gathering reliable contextual data.

Making PAL ubiquitous. As discussed in Section 3, we targeted a mobile/wearable solution for PAL. Our intuition was that the mobile phone would likely be with an individual most of the time (at least during working hours, perhaps also at home). Of the participants in the laboratory study who owned a mobile phone, all but one was carrying it upon arrival for the study. Furthermore, in 79% of the diary entries queried, the participant's mobile phone was on her or within reach.

The results of both studies demonstrated the need for and appropriateness of this service to be wearable, as opposed to environmental. The argument can be made that an audio buffering service in the environment might be preferable for a variety of reasons, including power concerns, better audio quality, and the convenience of users not needing to wear a device. Every participant reported, however, that there are times when it would not be possible for the service to be environmental. Every participant who recorded any entries recorded at least one at a public place or outdoors, where environmental solutions would be difficult. Participants also expressed control concerns about an environmental version of PAL versus a wearable solution. One participant noted, "[I would] rather have the control of it being on my person."

While advocating a wearable solution, however, participants were not interested in a completely separate device but instead as a "value added features" to the mobile phone already owned and carried. Although this may seem obvious in retrospect, it implies the fairly strict requirements that PAL must run unattended on the mobile handset, without recharging for at least a day, and it must not interfere with the call functions of the phone. These requirements are met by our currently deployed prototype, resulting in an arguably ubiquitous service.

Making PAL usable. Our final prototype provides asymmetric backward/forward skip features over the recording, with default values of 10 and 5 seconds, respectively. While most participants of the laboratory study liked these defaults, the values can be adjusted, and anecdotal experience shows that individuals do optimize them. We did not observe effective use of fast forward or rewind skimming features during the laboratory study. Considering the limited capabilities of the handset, we opted to support earmarks instead. The user can set earmarks and can use the backward/forward skip buttons to traverse these earmarks or to simply navigate without using them.

One issue identified in the laboratory study related to the mapping of the pair of navigation buttons: there is no "natural association" between the buttons and backward and forward navigation. This issue is exacerbated by the variety of ways the handset can be mounted on a belt or carried in the pocket or purse. We opted for a "never-wrong" mapping. When transitioning from record to playback, the only possible direction of navigation is backward in time. Therefore, whichever navigation button the user first presses is mapped to backward navigation; the other button is used for forward navigation. Once recording resumes, the previous mapping is cancelled.

Making PAL socially and legally acceptable. We do not endorse the common opinion that people necessarily must adapt to technological evolution by changing their social expectations. However, a case could be made that PAL does not impinge on constitutional rights and that, in the long term, practice could show the harmlessness of this application, granted specific guarantees, namely, small recording radius, short buffer length and some form of notification to the conversation partners. We would like to stress that it is not in the scope of this paper to provide conclusive legal opinions – a task best left to courts and DPAs. Our purpose is to provide a balanced, if necessarily concise, overview of PAL's social and legal impact.

A number of different stakeholders can be identified with regards to PAL; we consider three: the user, conversation partners and unrelated third parties. Considering the third category, diary results indicate that 69% of the entries related to recordings in public or semi-public spaces, and 44% stated that other, unrelated, people were present. These figures support our concern with third-party privacy, which contrasts with the fact that the vast majority of our participants were neither preoccupied with a third party's privacy nor with that of the conversation partner. These observations are particularly interesting because they diverge from legislation in force. ECPA does prohibit capturing a third party's conversation when the owner of the device is not part of that conversation and the conversation takes place with reasonable expectation that it is not being intercepted (*e.g.*, non-public space). On the other hand, it must be noted that the perceptual properties of sound might not grant *constitutional* basis (in the US) for an expectation of privacy in public space, as suggested among others by numerous cases adopting the "plain view" rule. This could allow adapting surveillance legislation to permit limited memory aid devices such as PAL.

Interface affordances and information retention policies greatly impact social acceptability. Altering the coverage of the microphone is an essential factor of a proportionality determination, as suggested by analogous DPA opinions involving personal uses of video surveillance (namely, outdoor camera units at home entrances) [2]. Likewise, DPAs have used retention time and deletion policies to evaluate the social impact of surveillance applications. Completely eliminating the risk of recording third parties' conversations is extremely difficult, given the characteristics of sound transmission, but the retention properties of this application do support the claim that PAL does not serve archival purposes, nor does it vastly facilitate surveillance, since the device is carried around by its user; if concealed or left unattended, the application arguably presents lower risks than traditional audio recorders.

In the relationship with conversation partners, informed consent is one fundamental tool of social action, embodied in privacy law. Its implementation presents though formidable technical and usability challenges. In our case, anecdotal evidence col-

lected during the deployment suggests that our participants have, over time, renounced to preventively explain or ask permission to use the service. At times participants turn off the device due to social pressure. Both observations support our previous findings from the diary study. This could hint at a gradual adaptation to the technology, and the adoption of appropriate social behavior, similarly to what is currently happening with camera phones.

Directive 95/46/EC exempts the personal use of information (*e.g.*, diary) from the informed consent requirements, and the figures reported above regarding control and usage seem to confirm that users view PAL as a preeminently personal application. When asked about objections by conversation partners, one participant answered “I wouldn’t care. It’s a tool for me.” Moreover, all-party informed consent would place an unreasonable burden on the user (a condition which may exempt from the consent requirement). Still, it is not guaranteed that this application would qualify as personal, nor that the Directive’s provisions, thought for textual diaries and address books would transfer in DPAs’ judgment to environmental recording. If not so, DPAs have expressed the need for explicit notification and consent. ECPA provides in the general case the “one-party consent” rule, in which informed consent by conversation partners is not necessary if the user of the recording device takes part in the conversation, without prejudice on the legality of the subsequent use of that information. ECPA acts only as a baseline, however, and many states have introduced various additional safeguards, such as two-party consent and notification cues such as “recorder beeps” (a useful, non-authoritative comparison of US state laws can be found in [12]).

Although we did not receive strong feedback from our participants requesting that PAL provide a notification cue while recording, in view of the above considerations, we decided to incorporate such function in the deployed handsets. When recording, the outer LED integrated in the round ornament on the phone shell (see Fig. 1) lights up red. During playback the light turns green. Although recording is usually associated with a red indicator, we are aware that people might not understand its meaning and that users could obviously conceal the LED as well as the recording device: the user remains ultimately responsible for abiding to the social contract and mores.

Concluding, the legality of PAL in parts of the US with stronger safeguards appears to be more problematic than in Europe because of the greater flexibility granted by EU law to DPA judgment. The lack of precedents and novelty of this recording without archiving do not allow us, however, to reach any definitive conclusion. In any case, characterizing PAL as a memory aid and *not* as a recording device appears to be the juncture through which any argument in favor of social and legal acceptability must flow.

6 Conclusions and Future Work

Based on controlled and field studies of use of a mobile audio-based memory aid, we conclude that not only is the service desirable for users, but also that its implementation on a mobile phone is possible and usable. Users can find the information needed in less time than they reported being willing to spend. They need this service at least once a week, and they are willing to wear a mobile phone at all times to have access to it. Our analysis shows that this application falls within a legal “grey area”, and that we cannot definitively assert or deny its legality. The interface and retention charac-

teristics of the application, along with observation of initial deployment suggest that the application might be socially acceptable. We have deployed PAL on the Motorola i730 platform and plan to report on a long-term study of the emergent uses PAL inspires and on the social contract and mores it influences.

Acknowledgements. Thanks to Motorola Corporation, and specifically Joe Dvorak of the iDEN group, for supplying the i730 handsets used for these studies. This work is supported in part by NSF ITR grant 0121661, the NSF GRFP, the Sam Nunn Security Program and Intel Corporation. All human subjects research conducted for this paper is covered under IRB approved protocols H03221 and H03217 at the Georgia Institute of Technology.

References

1. Arons, B.: SpeechSkimmer: Interactively Skimming Recorded. In: Proceedings of the Sixth Annual Symposium on User Interface Software and Technology (UIST '93). ACM Press, New York (1993) 187–196
2. British Institute of International and Comparative Law: The implementation of Directive 95/46/EC to the Processing of Sound and Image Data. (2003) http://www.europa.eu.int
3. Czerwinski M., Horvitz E., Wilhite S.: A diary study of task switching and interruptions. In: Proceedings of ACM Conference on Human Factors in Computing Systems (CHI 2004). ACM Press, New York (2004)
4. Deitz, P., Yerazunis, W.: Real-Time Audio Buffering for Telephone Applications. In: Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST 2001). ACM Press, New York (2001) 193–194
5. European Commission Art.29 Working Party: Opinion 4/2004 on the Processing of personal data by means of video surveillance, 11750/02/ENWP89 (2004) http://www.europa.eu.int
6. European Union: Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (Data protection directive)
7. European Union: Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)
8. Harrison, S., Dourish, P.: Re-place-ing space: The roles of space and place in collaborative systems. In: Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (CSCW'96). ACM Press, New York (1996) 67–76
9. Hindus, D., Schmandt, C.: Ubiquitous Audio: Capturing Spontaneous Collaboration. In: Proceedings of the 1992 ACM Conference on Computer Supported Cooperative Work (CSCW'92). ACM Press, New York (1992) 210–217
10. *Katz v. United States,* 389 U.S. 347; 88 S. Ct. 507 (1967)
11. *Kyllo v. United States* 121 S. Ct. 2038; 150 L. Ed. 2d 94; 2001 U.S. LEXIS 4487 (2001).
12. Radio-Television News Directors Association: Hidden Cameras, Hidden Microphones: At the Crossroads of Journalism, Ethics and the Law (1998) http://www.rtndf.org/
13. Tran Q.T., Mynatt, E.D.: What Was I Cooking? Towards Déjà Vu Displays of Everyday Memory. GVU Technical Report 03-33. Georgia Institute of Technology (2003)
14. United States Electronic Communications Privacy Act of 1986, 18 USC 2510

A Study on Gestural Interaction with a 3D Audio Display

Georgios Marentakis and Stephen A. Brewster

Glasgow Interactive Systems Group
Department of Computing Science
University of Glasgow
Glasgow, G12 8QQ, UK
{georgios,stephen}@dcs.gla.ac.uk
www.audioclouds.org

Abstract. The study reported here investigates the design and evaluation of a gesture-controlled, spatially-arranged auditory user interface for a mobile computer. Such an interface may provide a solution to the problem of limited screen space in handheld devices and lead to an effective interface for mobile/eyes-free computing. To better understand how we might design such an interface, our study compared three potential interaction techniques: head nodding, pointing with a finger and pointing on a touch tablet to select an item in exocentric 3D audio space. The effects of sound direction and interaction technique on the browsing and selection process were analyzed. An estimate of the size of the minimum selection area that would allow efficient 3D sound selection is provided for each interaction technique. Browsing using the touch screen was found to be more accurate than the other two techniques, but participants found it significantly harder to use.

1 Introduction

Designing a user interface for a handheld device to be used on the move is a challenging task. The lack of screen space for information display in combination with the disturbances incurred by walking makes most of the techniques that are used in desktop user interface design problematic. Anyone who has tried to read a piece of text on a handheld computer while sitting in a taxi or to target a menu item while walking can verify that this task is a difficult one.

We are taking an alternative approach to interface design for mobile devices by creating multimodal interfaces based on sound and gestures. Multimodal interfaces allow the user to use multiple senses to interact with a mobile computer. It is an objective of our work to use the human senses so that they act in a complementary way to each other. No sense can replace all of the others and each can outperform the rest for certain tasks. For example, listening to text is much more efficient than reading it when walking but on the other hand performing corrections and editing the result can be more efficiently done using the visual sense.

The study reported here examines the potential of designing an interface based on the auditory sense for information display and the use of gestures for control. Moreover, three-dimensional (3D) sound is used as it enables better separation between multiple sound sources and increases the information content of an audio display. It

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 180–191, 2004.

also allows the spatial nature of the audio space to be used, which we hope will be as beneficial as the spatial display of information in a Graphical User Interface (GUI).

The spatial aspects of our auditory sense have been little explored in human-computer interaction. The ability of the auditory system to separate and apply focus to a sound source in the presence of others (commonly known as the 'Cocktail Party' effect [1]) is very helpful for interface design. It implies that simultaneous streams of information can be presented, with users choosing to focus on what is most important (just as occurs visually in GUIs). This phenomenon is greatly enhanced if the sources are spatially separated and thus suggesting the use of three-dimensional sound in auditory user interface design. Other interesting audition properties include omni-directionality and persistence.

Gestures have the potential to be effective input techniques when used on the move because they do not require visual attention (as do most current mobile input techniques such as pens or soft keyboards). Our kinaesthetic system allows us to know the position of our limbs and body even though we cannot see them. This means that for a mobile application the user would not need to look at his/her hands to provide input, visual attention could remain elsewhere, for example on navigating the environment.

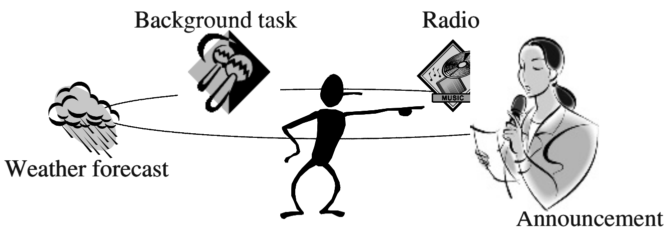


Fig. 1. Example of a gesture controlled 3D auditory user interface. A range of different audio sources are presented around a listener and they can be selected using a gesture.

As can be seen in the Figure 1, we are planning to build a 3D audio system where the user will be able to monitor a number of tasks simultaneously, discriminating between foreground and background ones and interacting with them using gestures. The user will hear a range of different sounds but will be able to tune in to the one that is most important, selecting items and interacting with them using gestures. The sound locations in this study are not truly three-dimensional. We place sounds on a plane around the user's head at the height of the ears to avoid problems related to elevation perception. This results in a 2.5D planar soundscape.

2 Previous Work on Auditory and Gestural Interfaces for Mobile Devices

Applications of audio in user interface design have been examined by many researchers. Gaver [8] introduced the notion of Auditory Icons in user interface design. Auditory Icons are based on the notion of everyday listening and they have been used in systems such as the SonicFinder and the ARKola system [15]. Blattner *et al.* [2] have proposed designing audio displays that are based on structured musical listening, re-

sulting in the notion of Earcons that have been examined and proved to be usable by Brewster [5].

The notion of an audio window system was introduced by Cohen and Ludwig [7]. Cohen also introduced the concept of using 3D sound to increase the auditory display space and proposed simple gestural interaction with 3D sound for input [6]. According to Cohen, sounds are positioned in the space around the user and a mapping between the sounds and the elements of the interface is performed. Users can subsequently interact with the sounds by pointing, pitching, catching and throwing them. By using these interaction techniques users can organize the system so that it suits their needs. Another idea developed by Cohen has been the an audio pointer as an aid in the cluttered audio space to assist localization and help the user disambiguate current position in relation to the position of the sounds in the display. The concept of 'filtears' has been also introduced by Cohen. According to this idea sounds slightly change as a result of filtering when being in different states such as selected, caught etc. This cue has been designed to assist the user in understanding the state of the display elements as he/she is interacting with them.

Another attempt to construct a system based around spatialised audio was Nomadic Radio by Sawhney and Schmandt [14]. It is targeted primarily at messaging. It is enabled with speech recognition and synthesis to allow the user to communicate and receive feedback from the system. It is also enabled with 3D audio to enhance simultaneous listening and conferencing. Another interesting issue about this application is the fact that it works based on loudspeakers mounted on the shoulders of the user and a directional microphone on the chest of the user; the user is able to listen to his/her real audio environment at the same time as when interacting with the system. The system also uses a space to time metaphor to position different messages around the user depending on the time of arrival. It works using a limited set of commands that can be recognized through the speech recognizer.

Brewster *et al.* [4] tested a three dimensional gesture controlled audio display on the move. They used an auditory pie menu centred on the head of the user and compared fixed to the world versus fixed to user sound presentation. They found that fixed to user sound presentation performs better in terms of time required to perform tasks as well as in terms of the walking speed the users could maintain. In another study by Pirhonen *et al.* [12] gestural control of a MP3 audio player was found to be faster and less demanding than the usual stylus based interaction when on the move.

Goose *et al.* [9] presented a system using 3D audio and earcons and text to speech for browsing the WWW. Finally, Savidis *et al.* [13] used a non-visual 3D audio environment to allow blind users to interact with standard GUIs. Different menu items were mapped to different locations around the user's head.

The ideas in the literature shape a framework for working with sound in a gesture controlled 3D audio display. Speech control has been used to control a 3D audio display, however it is known to require a silent environment to operate, users to be able to remember the command repertoire and can be indiscrete. Gesture control seems like a more feasible solution for systems to be used on the move and in a social context. Cohen as well as other researchers, have proposed designs for 3D audio interface development. However, with the exemption of [4], no formal evaluation of these ideas was done. We believe that given the ambiguity that can occur in such interfaces, further empirical research is necessary to allow us to design 3D audio interfaces in a formal way.

3 Three-Dimensional Audio Issues and Definitions

Designing a user interface based on 3D sound and controlling it by gestures poses a number of questions that must be answered before successful interfaces can be created. It is the case that when asking people to locate a sound, there is a certain extent of ambiguity in their answers. This ambiguity, called Localization Blur [3], has been measured for users listening to sounds from different locations in space and has been shown to be bounded (for a full review see [3]). As found in Blauert [3] localization blur can range from ±3.6° in the frontal direction, ±10° on the left/right directions and ±5.5° to the back of a listener under well controlled conditions. Localization blur also depends on the position of the sound source and the spectral content of the source. Virtual sound positioning using headphones is realized using HRTF filtering [3]. Head Related Transfer Functions (HRTF) are functions that capture the frequency response of the path between a sound source and the listener's tympanic membrane. These functions are estimated experimentally usually using a dummy head and torso. By filtering a sound signal using these functions it is possible to apply to it directional characteristics. However, problems related to non-individualized HRTF's (using a set of filters not created from your own ears) and HRTF interpolation and reproduction reliability affects the quality of the result so that performance is commonly poorer than for real-world listening.

In the light of these facts, it is interesting to try to define what we mean by asking a person to interact with a spatially positioned sound source, utilizing cues such as the source's direction. It is necessary to associate a certain area of the display to each of its elements. This mapping is not obvious as it is in graphical displays, since a person cannot judge exactly where the sound source is located or what its dimensions are. For example, consider the setup where non-overlapping sounds are presented around the user in the horizontal plane. In this case, we could map an angle interval to each display element. Any type of interaction that occurs in this area could be mapped to a specific display element positioned in the centre of this angle interval. By estimating this interval a design principle is obtained that can be used to partition the audio space. The estimation of such quantities can be problematic though, due to the unfamiliarity of many users with the sound localization task as well as with virtual 3D sound environments. Both when using real sound sources and when using virtual ones untrained subjects respond with great variation to questions related to the direction of a sound source.

Localization accuracy can also be improved by using feedback. Feedback could help in assisting the whole localization procedure by guiding the user towards the source and by reassuring the user that he/she is on the target area, thus making the selection process more effective. It could help overcome the poorer localization that occurs with virtual 3D sound to allow it to be used effectively in a user interface.

Two design techniques are positioning the sound sources egocentric versus exocentric. Egocentric or fixed to the listener sources, can be localized faster but less accurately, due to the absence of active listening. By active listening we refer to the process of disambiguating sound direction by small head movements. Active listening enhances localization accuracy but results in computationally intensive updating of the sound source positions (which may be a problem in a lower-powered mobile device) as well as in increasing the time required for a person to localize a sound stimulus. This is because the process of active listening involves moving and converging

towards the target sound using the information provided by the updated sound scene. There is a trade-off in localization accuracy and time required to make a selection when deciding between fixed to the listener versus fixed to the world sound sources. We chose the better accuracy of exocentric or fixed to the world over egocentric to overcome the limitations of the non-individualized HRTF's we used.

A key issue in 3D audio design is the number of sources that can be presented simultaneously. It has been shown that human performance degrades as the number of audio display elements increases [11] when sounds stem from the same point in the display. Spatial separation, however, forms a basic dimension in auditory stream segregation and thus can possibly increase the number of sources users can deal with. The study we present here uses just one sound source as we wanted to gain an idea of selection angles in the simplest case, before we move on to more sophisticated sound designs later in our research.

To handle the ambiguity in the aforementioned tasks, we decided to use adaptive psychophysical methods. Adaptive methods are characterized by the fact that a stimulus is adjusted depending on the course of an experiment. They result in measures of performance on psychophysical tasks as a function of stimulus strength or other characteristics. The result constitutes what is called a psychometric function [10]. The psychometric function provides fundamental data for psychophysics, with abscissa being the stimulus magnitude and the ordinate measuring the subjective response. One commonly used psychophysical method is the Up-Down method. Up - Down procedures work by setting the stimulus to a certain level at the beginning of an experiment and then decreasing or increasing the stimulus based on the observation of a specific pattern in the subject's response. The phenomenon that occurs when the direction of stimulus change is reversed is called a reversal. Up-Down methods that decrease the stimulus after a valid answer and increase stimulus after an invalid answer converge to the 50% point of the associated psychometric function. A point of this function that corresponds to 50% would imply that at this stimulus level, 50% of the answers would be expected to be 'valid'. By altering the rule of stimulus change, different points of the psychometric function can be estimated. However, full sampling of the function is often impossible due to the large number of experimental trials required.

4 Experiment

An experiment was designed to answer some fundamental questions about the design of audio and gestural interfaces, in particular: what is the minimum display area needed for the effective selection of a sound source, and what selection technique is the most accurate. We estimated the angle interval that would result in 67% of a user's selections being on target. To do this we used an adaptive psychophysical method, more specifically a two-down one-up method (for a review of adaptive psychophysical methods see [10]). We investigated three different browsing and selection gestures that could be used by users to find items in a soundscape and select them. We used head/hand tracking to update the soundscape in real time to improve localization accuracy.

The three browsing gestures were: browsing with the head, browsing with the hand or browsing using a touch tablet. These gestures differ with respect to how common

they are in everyday life. The first is the normal way humans perform active listening, with the position of the sound being updated as the user's head moves, so should be very easy to perform. The second is more like holding a microphone and moving it around a space to listen for sounds. The location of the sounds in the display is updated based on the direction of the right index finger. Direction is inferred by a 2D vector defined by the position of the head and the position of the index finger of the user. The third gesture can be thought as an extreme, in the sense that it cannot be mapped to a real world case. The user moves a stylus around the circumference of a circle on a tablet (the centre of the tablet marks the centre of the audio space) and the position of the sound source is determined by the stylus direction with respect to the centre of the tablet. In early pilot testing this type of sound positioning proved to be confusing if a user was to start a selection from the lower hemisphere. This was due to the fact that sounds moved as if the participant was looking backwards, although the participant was actually looking forwards. For this reason, we decided to reverse left and right in case the user began browsing in the lower hemisphere. By doing this, the optimal path to the next sound could be found by always moving on the circle towards the direction in which the sound cue was perceived to be stronger.

The selection gestures were: nodding with the head, moving the index finger as if clicking a non-existent mouse button, and clicking a button available on the side of the stylus to indicate selection. In this experiment, three combinations of the above were examined: browsing with the head and selecting by nodding, browsing with the hand and selecting by gesturing with the index finger, and browsing with the pen on the tablet and selecting by clicking.

4.1 Sound Design and Apparatus

The aim of the experiment was to look at how the minimum angle interval that allows efficient selection of an audio source varies with respect to direction of sound event and interaction technique used. We used a single target sound placed in one of eight locations around the users head (every 45° starting from 0° in front of the user's nose) at a distance of two meters. This stimulus was a 0.9 second broadband electronic synthesizer sound, repeated every 1.2 seconds.

We used very simple audio feedback to indicate that the user was within the target region and could select the sound source. This was a short percussive sound that was played repeatedly while the user was 'on target' (i.e. within the current selection region) to assist each user in localizing the sound. This was played from the direction of the target sound. Sounds were played via headphones and spatially positioned in real time using the HRTF filtering implementation from Microsoft's DirectX 9 API. Sound positions were updated every 50msec.

To perform gesture recognition and finger tracking we used a Polhemus Fastrack to get position and orientation data, and two sensors (see Figure 2). One sensor was mounted on top of the headphones to determine head orientation and allow us to recognize the nod gestures. A second sensor was mounted on top of the index finger to determine the orientation of the hand relative to the head and to recognize the clicking gesture in the hand condition. A Wacom tablet was used for the tablet condition. We determined nodding and clicking by calculating velocity from the position data.

Fig. 2. A participant making a selection in the hand pointing condition.

4.2 Experimental Design and Procedure

The experiment used a two-factor within-subjects design with each participant using each of the three interaction techniques in a counterbalanced order. There were two independent variables: sound location (eight different levels) and interaction technique (three levels). The dependent variables were deviation angle from target and effective selection angle. Participants were also asked to rate the three interaction used for browsing and selecting on a scale from one to ten with respect to how comfortable and how easy to use they found them. Our hypotheses were that the effective selection angle would be affected by interaction technique, with no effect of location because participants always faced the targets when selecting them.

Twelve participants took part: five females and seven males with ages ranging from 19 to 30.

The participant's task was to browse the soundscape until the sound was in front and then select the target sound using the interaction techniques described. The target sound repeated until the participant performed a selection. Upon selection, the stimulus was presented in a different location randomly out of the set of available positions. The whole process was repeated until all up-down methods for each position converged. According to the up-down rule the effective selection angle was varied between trials; it was reduced after two on-target selections and increased after one off target selection. The step was initially 2° but was halved to 1° after the third reversal occurred. It should be noted that participants were unaware of this process; they were instructed to perform selections based only on audio feedback and localization cues.

The experiment lasted approximately one hour. Participants stood wearing the headphones and tracker. They could turn around and move/point as they wished and were given a rest after each condition. The experiment could not be conducted in a fully mobile way with users walking (as in previous studies such as [4]) due to the tracking technology needed for gesture recognition – participants had to stay within range of the Polhemus receiver. The results may therefore be different if the techniques were used in a fully mobile setting, but they will indicate if any of them are usable and should be taken further. Participants were trained for a short period before being tested in each condition to ensure they were familiar with the interaction techniques. They performed eight selections before embarking on the experiment. Prior to

testing, participants' localisation skills were checked to rule out hearing problems and to familiarise them with the sound signal they would hear. During this 3D sound training, participants were asked to indicate verbally the direction they had perceived the sound source was coming from. The experimenter subsequently corrected them in case they were wrong and tried to direct their attention to the relevant cues.

5 Results

The up-down method was expected to converge on the point of the associated psychometric function where 67% of the selections would be on target. To estimate this point we averaged the angle intervals as these were updated by the up-down rule. Averaging included only the angle intervals that occurred after the second reversal.

A 3x8 two factor ANOVA was performed to examine whether sound location and interaction technique affected the effective selection angle. Sound location was not found to have a significant main effect ($F(2.314, 77) = 2.241$, $p = 0.121$). However, there was a significant main effect for interaction technique ($F(2, 22) = 10.777$, $p = 0.001$). There was no interaction between location and technique. Pair-wise comparisons using Bonferroni confidence interval adjustments showed that the tablet condition was significantly more accurate than the other two techniques, but no significant differences were found between the hand and head. Figure 3 shows the mean effective angle intervals for the three interaction techniques with respect to direction of the sound. These results define the one side interval around a source. To give an example of how these data could be applied, if an exocentric 3D audio user interface (enabled with active listening) using audio feedback and controlled by a stylus on a touch tablet, was developed, the designer should allow at least 4° on *each side* of a sound positioned at 90° relative to the front of the user so that a user would be able to select the sound effectively.

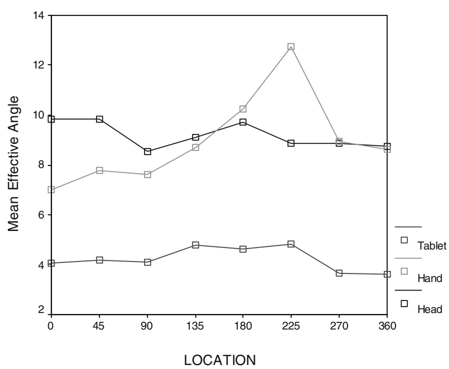


Fig. 3. Effective selection angle for each sound direction.

The deviations of the users' selections from target were also analyzed. Ninety measurements for all different directions were analyzed. A 3x8 two factor ANOVA showed a significant main effect for interaction technique ($F(2,192) = 7.463$, $p = 0.001$). Direction also had a significant main effect ($F(7,672) = 7.987$, $p = 0.001$).

There was a significant interaction between technique and direction (F(14,1344) = 7.996, p = 0.001). Pair-wise comparisons using Bonferroni confidence interval adjustments showed that the tablet condition was significantly better than the others, but there was no significant difference between head and hand. With respect to the direction of the sound event, direction 225° was significantly different from direction s 0°, 45°, 90°, 135°, 270°, 315° and direction 180° was different from 45°, 270°, 315°. Figure 4 illustrates mean deviation from target and its standard deviation.

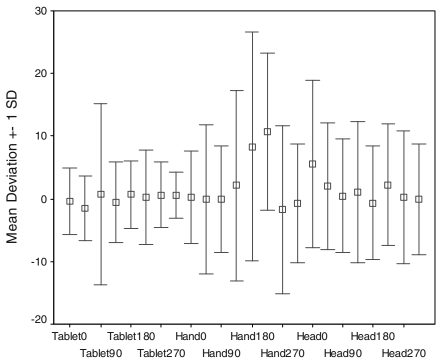


Fig. 4. Mean deviation from target versus sound direction.

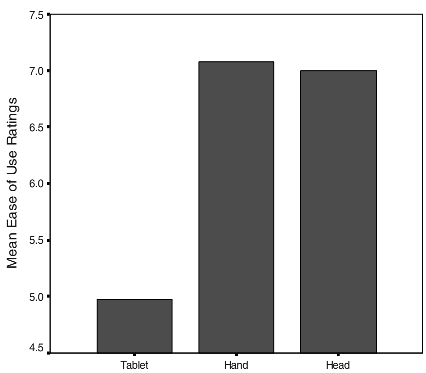


Fig. 5. Mean ease of use ratings for each interaction technique.

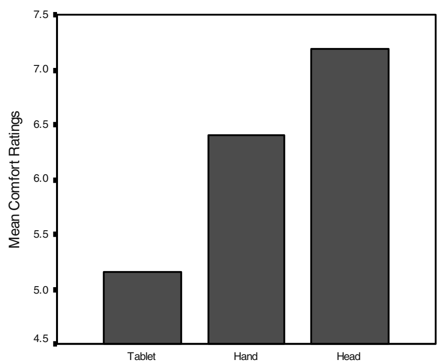


Fig. 6. Mean Comfort ratings for each interaction technique.

As mentioned, each participant was asked to rate each of the interaction methods in terms of how easy and how comfortable he/she found them to be, on a scale from 1 to 10. Figure 5 shows the means of the results for ease of use. A statistical analysis of variance showed interaction method to be a significant factor (F(36) = 7.386, p = 0.002). Bonferroni t-tests verified mouse to be significantly harder to use, but showed no statistical difference between hand and head.

A similar analysis on how comfortable the use of the three devices was showed no significant difference between devices. Figure 6 shows comfort means for the three interaction methods. It should be noted that participants have performed a large num-

ber of selection using the three interaction methods to allow the up-down methods to converge in the three different conditions and the eight different sound positions. In that sense, when observing the graphs, absolute values should be taken into account carefully. However, the ratings of the three devices relative to each other can be used to infer how they are ordered relative to each other with respect to ease of use and comfort.

6 Discussion

The results of the study showed that interaction with a 3D audio source can be done effectively in the presence of localization feedback. They also showed that novel methods of browsing can be as effective (and even more effective) in locating sounds than 'natural' ones in the presence of feedback. Users were able to perform active listening using the tablet and the hand without any particular difficulty. It was also surprising that they could do the active listening operation more accurately when using the tablet than when using their heads. This can be explained in terms of the resolution that the three different mechanisms provide. For example, a stylus controlled touch tablet provides a much better minimum possible displacement compared to the head or the hand of a person. By constructing histograms of deviation data we verified that the results of the up-down procedure would indeed allow 67% percent of the selections to be on target. It should be mentioned, however, that more reversals would result in having more accurate results. This was not possible to do since we tried to maintain a within-subjects design and keep the experiment duration in the order of one hour to avoid effects caused by fatigue. Effective selection angles are likely to reduce with practice and improved feedback design.

When considering the three interaction methods, one would not expect the direction of sound to be a significant factor in the results of this study. This is due to the active listening operation; that is, users selected a sound when it was in front of them. This was verified in the effective angle case where no location was found to be a significant factor. However, in the deviation analysis, certain angles were significantly different from others. This was mostly in the direction of 225° degrees. The reason for this difference can be described by the mechanics of the browsing and selection modalities. A closer look at the graphs reveals the technique that caused this difference was browsing by hand. As was observed during testing, some right-handed participants found it difficult to point to that location, if they had not turned their bodies first (they had to reach around their body causing them to stretch, reducing the accuracy of their selections). A significant number of participants indeed tried to point without turning their bodies, a result that influenced the accuracy of the browsing and selection processes.

By analyzing how the ease of use ratings are ordered, we see that users find browsing the sound space to be equally easy either using the head or using the hand. The touch tablet however, although more accurate, was not rated highly. This can be associated with the unnaturalness of the browsing process. In the other two cases, participants used a natural process for browsing the space, such as moving their heads or simulated one by moving their hand in a synchronous way with their head.

When considering the effective angles, we can observe that if accuracy was the only factor to be taken into account, an audio user interface could be constructed hav-

ing all eight sounds locations, and possibly more. Our next study will investigate the presentation of multiple sounds and the design of a more sophisticated soundscape such as would be needed for a real application of a wearable device based around 3D sound and gestures. If studies show that listeners cannot use sounds from eight locations then we can increase the selection angles for our sound sources which will further increase selection accuracy.

7 Conclusions

In this paper a study on gestural interaction with a sound source in the presence of feedback was presented. Three different gestures for browsing and selecting in a 3D soundscape were examined and their effectiveness in terms of accuracy was assessed. Browsing and selecting using a touch tablet proved to be more accurate than using a hand or a head gesture. However, browsing and selecting using the hand or the head were found to be easier and more comfortable by the users. Effective selection angles that would allow efficient selection were estimated for each interaction technique and on 8 sound locations around the user using an adaptive psychophysical method. The results show that these different interaction techniques were effective and could be used in a future mobile device to provide a flexible, eyes free way to interact with a system.

Acknowledgements. This study was supported by the EPSRC funded Audioclouds project grant number GR/R98105.

References

1. Arons, B., *A Review of the Cocktail Party Effect.* Journal of the American Voice I/O Society, 1992. **12**: p. 35-50.
2. Blattner, M. M., Sumikawa, D. A., and Greenberg, R. M., *Earcons and Icons: Their Structure and Common Design Principles.* Human-Computer Interaction, 1989. **4**(1): p. 11-44.
3. Blauert, J., *Spatial Hearing: The psychophysics of human sound localization*. 1999: The MIT Press.
4. Brewster , S., Lumsden , J., Bell, M., Hall, M., and Tasker, S. *Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices*. In *ACM CHI*, 2003. Fort Lauderdale, FL: ACM Press, Addison-Wesley. p. 463-480
5. Brewster, S. A., *The design of sonically-enhanced widgets.* Interacting with Computers, 1998. **11(2)**: p. 211-235.
6. Cohen , M., *Throwing, pitching and catching sound: audio windowing models and modes.* International Journal of Man - Machine Studies (1993), 1993. **39**: p. 269 - 304.
7. Cohen , M. and Ludwig , L., *Multidimensional Audio Window Management.* Internaltional Journal of Man - Machine Studies, 1991. **34**: p. 319-336.
8. Gaver , W. W., *The SonicFinder: An Interface that uses Auditory Icons.* Human-Computer Interaction, 1989. **4**: p. 67-94.
9. Goose, S. and Moller, C. *A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure*. In *7th ACM international conference on Multimedia*, 1999. Orlando, Florida, United States: ACM Press. p. 363 - 371

10. Leek, M. R., *Adaptive procedures in psychophysical research.* Journal of Perception & Psychophysics, 2001. **63**(8): p. 1279-1292.
11. McGookin, D. K. and Brewster, S. A. *An Investigation into the Identification of Concurrently Presented Earcons*. in *ICAD 2003*. (Boston, MA). p. 42-46
12. Pirhonen , A., Brewster , S., and Holguin, C. *Gestural and audio metaphors as a means of control for mobile devices*. In *ACM CHI*, 2002. Minneapolis, Minnesota, USA: ACM Press New York, NY, USA. p. 291-298
13. Savidis, A., Stephanidis, C., Korte, A., Rispien, K., and Fellbaum, C. *A generic direct-manipulation 3D auditory environment for hierarchical navigation in non-visual interaction.* In *ACM ASSETS '96*, 1996. Vancouver, Canada, 1996: ACM Press. p. 117-123
14. Sawhney, N. and Schmandt, C., *Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments.* ACM Transactions on Computer-Human Interaction, 2000. **7**(3): p. 353-383.
15. William W. Gaver, Randall B. Smith, and O'Shea, T. *Effective Sounds in Complex Systems: The ARKOLA Simulation*. in *Conference on Human Factors in Computing Systems archive.* In *ACMCHI.* 1991. New Orleans, Louisiana, United States: ACM Press New York, NY, USA. p. 85 - 90

A Study of Application and Device Effects Between a WAP Phone and a Palm PDA

Jiraporn Buranatrived and Paul Vickers

Northumbria University, School of Informatics,
Pandon Building, Camden Street, Newcastle upon Tyne, NE2 1XE, UK
paul.vickers@northumbria.ac.uk
http://www.paulvickers.com/northumbria

Abstract. Technologies like Java 2 Micro Edition and Microsoft's .NET framework allow applications to be developed and deployed across a range of mobile devices without having to significantly change the source code. However, mobile devices have very different interfaces and capabilities and it is not clear whether these generic deployment technologies adversely affect the usability of applications by ignoring individual device characteristics. This paper describes an experiment that aimed to see whether users of two applications written with J2ME and deployed on two devices experienced any differences in the usage of the applications on the different devices. Our findings indicate that usability can be maintained through multi-platform deployment, but that there are may also be usability advantages if the specific interaction paradigms of different mobile platforms are taken into account. This would require means of separating not just the interface from the functionality, but also the interface functionality from the interface data.

1 Introduction

Mobile computing is a fast-growing industry. Mobile devices are starting to replace older forms of communication and computation and developers are faced with new issues that need to be addressed throughout the development process [1]. Compared with desktop computers they have a number of substantial limitations (mainly associated with their memory, processing power, and their interfaces which are typically less sophisticated and relatively small). The emergence of wireless devices and mobile networks has opened up new business opportunities as e-commerce now extends into the mobile realm to become m-commerce (mobile commerce). To exploit the technical opportunities that mobile computing offers (such as instant connectivity, localization and the capability to receive information and conduct transactions anywhere, at any time, in a real-time environment) companies must develop effective and efficient applications with friendly and usable interfaces. Designing for mobility, a dispersed and widespread population, limited input and output capabilities, and supporting increased multitasking with more interruptions is a challenge that is coming to the fore [2]. Success will be affected by finding the right mix of applications that fit within constraints of limited screen size, memory, and processing power. Good interface design requires more than just squeezing information into a little screen.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 192–203, 2004.

There are several factors involved in building a successful m-commerce venture including security, networking technologies, and usability. In fact, as Table 1 shows, poor usability is rated second only to fraud concerns as an obstacle to consumer uptake of m-commerce.

Table 1. Obstacles to Consumer Adoption of m-Commerce [3]

Obstacle	Phones	PDAs
Credit card security concerns	52%	47%
Fear of 'klunky' user experience	**35%**	**31%**
Don't understand how it would work	16%	16%
Other	11%	13%
Never heard of it before	10%	12%

Success depends on finding the right mix of applications that fit within the constraints such as limited screen size, memory, and processing power. In the case of interface design, it requires more than just squeezing information into a tight little GUI. Designing a user interface that is successful within the constraints of mobile devices is, therefore, an interesting challenge. Usability can be viewed as having three broad dimensions: efficiency, effectiveness, and user satisfaction. Good usability is critical to attract and retain users but current mobile devices, with their small keyboards and displays pose interesting challenges to the mobile-interface designer.

M-commerce applications can be developed in various ways. A range of software and technologies are available to support this type of application, the two main ones today being Microsoft's .NET framework and Java (in form of J2ME—the Java 2 Micro Edition). There are different types of mobile device capable of supporting m-commerce, such as mobile phones and pocket computers or personal digital assistants (PDA). Each device (even within its own class) has its own unique user interface. This means there are likely to be different usability issues for each application running across different devices. By using a technology like J2ME the same functionality can be deployed across multiple platforms without the need for code rewrites. However, such an approach would not allow the application to be tailored to suit the individual interface capabilities of particular devices. In the interests of cost and efficiency it would be in developers' interest to use standardized deployments. Therefore, in an attempt to see whether just such a unified approach can result in applications being successfully deployed on different platforms without significant negative impact on usability we carried out an experiment in which two different types of m-commerce application were deployed on a mobile phone and a Palm OS PDA.

J2ME was chosen for the development platform because it is supported by most mobile phones and PDAs. Java technology-based architecture for m-commerce consists of four main tiers: back end tier, middleware tier, web tier, and client tier. The back end (or legacy) tier supports servers and mainframes running databases. The middleware tier is the connection buffer between the back end and web tiers. Enterprise Java Beans™ can be used to implement solutions in this tier. The web tier is a web server hosting JavaServer™ Pages, servlets, and Java Beans. The client tier is where J2ME is implemented on mobile devices.

2 Multi-platform Deployment

Chittaro and Dal Cin [4] discovered significant differences in the way navigation and item selection techniques affect interaction on a single mobile phone platform. Because there are several types of mobile device capable of supporting m-commerce, each with its own user interface, there will be different usability issues for an application running on different devices. Thus, we wished to see if we could implement common applications on different devices and yet maintain a broadly equivalent user experience.

Two prototype applications were built using J2ME: one to simulate mobile stock broking and the other to simulate the on-line purchasing of cinema tickets. J2ME allows applications to be compiled for both a Palm OS computer and a mobile phone without changing any of the code. There are obvious advantages to being able to write an application once and deploy it across different platforms. However, the benefits of this approach would be lessened if the usability of the application varied across the different devices, hence the reason for this study.

2.1 Movie Ticket Purchasing

Chittaro and Dal Cin [4] used a movie ticket purchasing scenario for their work. Movie ticket purchasing is a customer-driven activity in which the user requests information from a server and responds accordingly. We developed a simple prototype system that displays a list of available movie titles, a list of cinemas based on the user's location and, for a chosen movie and cinema, the list of showing times. Users select their seat position and specify the type and quantity of tickets required (e.g. adult, student, child, etc).

Fig. 1. Ticketing application showing the seat selection screen on phone and PDA

2.2 Stock Broking

Mobile Broking is a mobile financial services application that is commercially important. Financial institutions are using this new service channel because it supports convenience, timeliness, and decision-making. Location-independent, real-time information about share prices and the potential to act upon them is of high value to stock traders. Receiving alerts about price-movements and order executions, checking quotes, buying and selling stocks and other financial instruments are the key functionalities of mobile broking.

Whilst stock broking supports the same transaction model as ticket purchasing, it also has a real-time event-driven aspect. Stock prices are served at regular intervals to the mobile device. Price thresholds are set for certain stocks and the device signals an alert when chosen stocks hit these thresholds. The user then decides whether to buy or sell the shares. The broking application has a number of screens allowing the user to sign in, monitor stock prices, choose stocks to buy and sell, and execute stock transactions. Fig. 2 shows the application running on the phone and the Palm.

Fig. 2. Mobile broking 'confirm' screen on phone and Palm

3 Experiment

To study the effects on usability of running the two applications on two different platforms an experiment was carried out. The target markets for m-commerce consumer services are teenagers (18 years and under), students (19-25 years old) and young business-people (25-36 years old) [5]. Therefore, to cover two of the three target markets above, sixteen participants (8 male, 8 female) were chosen within the age range of 19-36 from an MSc Computing course. Moreover, this age group is more likely to be familiar with mobile phones, PDAs and mobile commerce applications. The subjects were required to carry out four tasks: one for each application/device pair (see Table 2). At the time of the study there were not sufficient Palms and phones

available so the experiment was run on emulators instead (with the PC keyboard disabled).

Table 2. Experimental tasks

Task		Description
1	Mobile broking on the mobile phone	Subjects were asked to monitor the prices of various stocks via a stock ticker and a series of alert messages that were triggered by the application when particular stocks hit pre-programmed price thresholds. The main task was to buy and sell a number of certain stocks when their prices were between pre-specified lower and upper limits.
2	Mobile broking on the PDA	This was the same as Task 1 except that it was performed on the PDA rather than the phone. The details of the stocks to be traded were also changed.
3	Ticket purchasing on the mobile phone	Subjects were required to buy a range of adult, student, and child tickets for specified films showing at specified times at certain cinemas. Subjects were also required to find out information about showing times of certain films.
4	Ticket purchasing on the PDA	This task was the same as task 3 except that it was performed on the PDA and with different film, time, and cinema ticket requirements.

To reduce the chance of any task order effects participants were randomly allocated to four evenly-sized groups comprising two male and two female subjects. Each group was allocated a different task order, thus:

Group 1: Task 1, Task 3, Task 2, Task 4
Group 2: Task 3, Task 1, Task 4, Task 2
Group 3: Task 2, Task 4, Task 1, Task 3
Group 4: Task 4, Task 2, Task 3, Task 1

Each subject's performance (such as time taken, error rate, etc.) was logged automatically by the two applications. Patterns of system usage, speed of completing task, rate of errors were traced in both applications and both devices. Correctness scores for subjects were calculated for each aspect of a task that was successfully completed. Subjects' workload for each task was measured using the NASA Task Load Index (TLX) method [6, 7]. TLX allows comparisons to be made between tasks in terms of the mental and physical demands experienced by the subjects.

The participants completed a short questionnaire about their past experience of using mobile phones and PDAs, and their past experience of buying stocks and shares and cinema tickets. Instructions were given telling the subjects what stocks and movie tickets they were to buy/sell.

Immediately following each task they completed a short two-part questionnaire. The first part asked a specific closed question that could only be answered by having used the application (for example, "what is the price of the stock SYSB?"). The second part asked for responses (rated on a five-point Likert scale) to nine statements

about the task (for example, “You would like to use this system to buy and sell stock”). Finally, subjects completed a TLX task-load rating sheet for the task.

We were hoping that J2ME would allow an application to be equally usable on either device.We specified the following hypotheses (with their corresponding null hypotheses) to investigate this:

- H1—There is a significant difference in task duration between the devices.
- H2—There is significant interaction between application and device on task duration.
- H3—There is significant difference in correctness scores between the applications.
- H4—There is significant difference in correctness scores between the devices.
- H5—There is significant interaction between mobile commerce application and mobile device on correctness scores.
- H6—There is significant difference in user satisfaction between the applications.
- H7—There is significant difference in user satisfaction between the devices.
- H7—There is significant interaction between application and device on user satisfaction.
- H8—There is significant difference in workload between the applications.
- H9—There is significant difference in workload between the devices.
- H10—There is significant interaction between application and device on workload.

4 Results

For each task there were four sets of results to be analysed: the time taken to complete the task, the correctness score, the questionnaire responses, and the subjects’ TLX workload assessments. The mean task results are shown in Table 3. Kurtosis and skewness analysis showed the raw data did not differ significantly from a normal distribution. Levene’s Test of Equality also revealed the data for each of the groups did not have significantly different variances. This meant that two-way ANOVAs could be used to look for effects between the applications and devices.

Table 3. Task results

	Stock broking		Ticket purchasing	
	Phone	PDA	Phone	PDA
Duration (sec.)	20	18	217	226
%Correct	94.53	93.75	92.19	95.70
Satisfaction	71.39	70.69	78.19	74.58
TLX score (workload)	48.44	43.12	25.44	29.44

4.1 Task Duration

Because more steps were involved in ticket purchasing (such as browsing film and cinema lists) it is not meaningful to compare task durations between the two applications. What is interesting here is whether the devices had any effect on the task duration. We can see from Table 3 that the difference in duration for the stock broking task is 2s. between the phone and the PDA and 9s. for the ticketing application. That is, the broking task took approximately 11% longer on the phone, whilst the ticket purchasing took approximately 5% longer on the PDA than on the phone. However, the two-way ANOVA (Table 4) shows no evidence for a main effect for the device or for an interaction effect between the device and the application. This means that the small differences in duration between the applications on the two platforms were not significant. Therefore, we reject hypotheses H1 and H2 and conclude that there was so significant difference in task duration between the two devices.

Table 4. ANOVA for task duration

Source	df	F	p
Device	1	0.052	0.820
Application * Device	1	0.114	0.737

In addition, using a keystroke level model (KLM) and Fitt's Law [8] we were able to calculate theoretical task durations for each application on the two devices. Normally, this KLM decomposes the task execution phase into five different physical operators (keystroking, pressing a mouse button, pointing or moving the mouse at a target, switching between mouse and keyboard, and drawing lines using the mouse), one mental operator (or mentally preparing for a physical action), and one system response operator. However, the application designs mean only two physical operators are used (pressing a button and navigating to a target) plus the mental and system operators. The time to locate the cursor target can be calculated using Fitt's Law. The time taken to hit a target is a function of the size of the target and the distance that has to be moved [9]. The common formula is:

$$\text{Movement time (sec)} = a + b \log2 (\text{distance/size} + 1) \quad (1)$$

where a and b are empirically determined constants [9]. Using recommended values for a and b suggested by Card, Moran and Newell [8]], the formula used becomes:

$$\text{Movement time (sec)} = 0.1 \log2 ((\text{distance/size}) + 0.5) \quad (2)$$

As the movement time depends on the position and size of the target, there are several movement times in this experiment. For instance, the time to move to the launch button in the broking task on the mobile phone is 0.55s. whereas on the PDA the calculated duration is 0.60s. The total durations for four experimental tasks were thus calculated and are shown in Table 5 along with the actual observed mean times.

Table 5. Theoretical vs. actual task times

	Stock broking		Ticket purchasing	
	Phone	PDA	Phone	PDA
Theoretical duration (sec.)	16	19	199	235
Mean actual duration (sec.)	20	18	217	226

We can see from Table 5 that the mean observed task times on the PDA are roughly 4-5% longer than the predicted times which suggests the model was a reasonably solid. Conversely, the predictions for the phone-based tasks are 25% longer than the observed value for the broking task and 10% longer for the ticketing task. That the subjects took less time on both phone-based tasks suggests that the KLM model doesn't translate as well to mobile phone interfaces as it does to PDAs. What these results suggest overall, though, is that the performance by the subjects was roughly in keeping with what one might expect from the KLM (the phone differences being noted) and so it makes sense to use the ANOVA to look for effects between the application and device times.

4.2 Task Correctness

Correctness was calculated by awarding a mark for satisfying each component of the transaction. For example, marks were awarded for choosing the correct cinema, film title, showing time, etc. Table 6 shows a non-significant difference in scores between both the application and the device. The device on which the tasks were performed also had no significant effect on the scores. Finally, there was no observed interaction effect between the application and device types and so we reject hypotheses H3, H4, and H5, and conclude that task correctness was not affected by either the application or the device.

Table 6. ANOVA for task correctness scores

Source	df	F	*p*
Application	1	0.006	0.939
Device	1	0.294	0.590
Application * Device	1	0.726	0.397

The high accuracy rates (above 90% - see Table 3 for details) suggest that the subjects understood the systems and the task requirements. Although the score differences were not statistically significant, the broking task on the phone and the ticket purchasing task on the PDA, whilst having similar means to their counterparts, had greater ranges of scores with some low outliers. It is possible that the higher mean for the PDA ticket purchasing arises from the difference in screen space on the two devices. Purchasing tickets required the user to look at lists of films, showing times,

cinemas etc. The smaller screen on the phone meant that users had to do more scrolling which may account for the higher error rate. Further experimentation could be done to explore this.

4.3 User Satisfaction

User satisfaction was measured for each task using a questionnaire. Percentage satisfaction scores were calculated for each task by deriving an overall rating for each subject's nine task-related responses and then computing the mean.

Table 7. User satisfaction for the applications and devices

	Broking % correct		Ticketing % correct	
Subject	PDA	Phone	PDA	Phone
1	22	20	32	35
2	32	39	34	40
3	28	21	31	36
4	40	38	40	35
5	34	34	32	40
6	33	35	32	35
7	34	30	36	35
8	30	25	33	34
9	41	40	41	37
10	33	34	36	36
11	35	26	32	35
12	31	33	32	35
13	30	38	35	36
14	33	36	38	37
15	16	27	22	19
16	37	38	31	38
Mean	31.81	32.13	33.56	35.19
Mean (%)	70.69	71.39	74.58	78.19

From Table 7 we can see that for both applications, the mobile phone had a higher mean satisfaction score than the PDA but that this difference is not significant (p>0.05 in all cases, see Table 8). Therefore, we reject hypotheses H6, H7, and H8 and conclude that user satisfaction was not affected by either the application or the device; that is, satisfaction was not significantly different across the devices or the applications.

It is slightly puzzling that the ticketing application on the phone had the highest satisfaction rating yet had a lower accuracy score than on the PDA. This may be because only three of the 16 subjects had prior experience of PDA usage whilst all subjects had used mobile phones before. Thus, the relative unfamiliarity of the PDA interface may have affected their opinions. This factor could also go some way to

Table 8. ANOVA for user satisfaction scores

Source	df	F	*p*
Application	1	3.053	0.086
Device	1	0.495	0.484
Application * Device	1	0.227	0.635

explaining why the mobile phone task durations were lower than predicted whilst the PDA task durations were slightly higher than was predicted by the model.

Analysis of the responses to the nine individual usability reveals that only statements 1 ("This system is easy to use") and 9 ("In general the system's response time was fast and you are satisfied with it") yielded results of statistical significance. For statement 1, an ANOVA suggests that ease of use was affected by the application (F=9.174, $p<0.01$) but not by the device (F=0.021, $p>0.05$) or by any interaction between device and application (F=0.187, $p>0.05$). Similarly, responses to statement 9 were significantly affected by the application (F=10.161, $p<0.01$) but not by the device (F=2.788, $p>0.05$) nor by the device/application interaction (F=0.023, $p>0.05$). This suggests that the device on which applications run has less impact on user acceptance than the applications themselves.

4.4 Workload

Using the NASA TLX method, subjects assessed their workload levels for the four tasks. TLX requires subjects to rate their experienced level of workload in five areas: mental demand, physical demand, temporal demand, effort, and performance. Subjects rank these five areas in terms of their perceived importance and contribution to the workload for each task. From these ratings and rankings an overall workload figure on a scale of 0-100 is calculated. The mean scores for each task are back in Table 3.

Table 9. ANOVA for TLX workload scores

Source	df	F	*p*
Application	1	15.278	0.000
Device	1	0.020	0.889
Application * Device	1	0.984	0.325

From Table 9 we see that the difference in workload between the two applications is highly significant ($p<0.01$). The devices themselves had no significant impact on the workload reported by the subjects. However, we do observe a slightly higher (though not statistically significant) workload rating for the mobile phone over the PDA in the broking application. This task required more data entry than the ticketing application (which had more list selection operations). It is possible that users might have found data entry harder to do with a phone keypad than with the PDA. (A single key on a phone keypad can be used to enter a digit, one of several upper and lower case characters, and one of a selection of punctuation/white space characters). Thus

we accept hypothesis H8 and conclude that the workload was higher for the stock broking application. However, we reject hypotheses H9 and H10 and conclude that the devices and the device/application interaction had no significant impact on subjective workload. Although the broking task involved higher workload, the fact that workload was not affected by the device differences suggests that the multi-platform deployment was successful.

5 Conclusions

Overall, the study indicated that the complexity and nature of the task itself appeared to be more important factors than the device on which they were run. This in turn suggests that provided a good task and interaction design is carried out, applications can be written once and rolled out across multiple platforms using technologies like J2ME. However, as the second study suggests uniformity of experience is probably achieved at the expense of the greater usability benefits that accrue from tailoring applications to exploit the interaction characteristics of individual devices.

The results of the study have provided interesting areas for further exploration. The results indicated higher error rates on mobile phone tasks that required scrolling through long lists. This aspect should be explored more formally to see what effect devices have on list scrolling tasks. Alternative representations could also be investigated to see how such data might be better presented on small screens. For instance, Brewster has provided some evidence that button size on a PDA can be reduced if auditory feedback is used [10] and Vickers and Alty have demonstrated that sound can be used to communicate quite sophisticated computing information [11, 12].

Technologies like J2ME allow developers to make economies by writing an application once and deploying it across multiple mobile platforms. All that is required is a device profile for each target platform. This has obvious advantages in terms of application consistency and development efficiency. However, it also means that each device renders the application with roughly the same interface widgets. This means that by ignoring the particular characteristics of individual devices, we might be missing out on significant usability and satisfaction improvements. The stylus-based interface allows much more direct manipulation than the phone interface. Homogenising the interface to work across all devices seems a retrograde step. Chittaro and Dal Cin [13] found significant differences in usability between the different ways of implementing navigation and item selection on a single WAP phone. It would seem advantageous if the application could make use of the best interface widget for the job on a given device.

This means that a way is needed not only of separating the core functionality from the interface but also of redefining the interface not as a set of widget/data pairs but as a description, for each task, of the task's I/O requirements. A device profile would then allow the application to determine, on a device-specific basis, how those I/O requirements should be rendered in terms of interface widgets. Preferably, the device profile could be locally configurable so that the user can override its defaults. This would also allow accessibility issues to be addressed as multi-modal-capable devices could use audio, graphics, and haptics in combination to suit the individual needs of

the user. Developing such 'plastic' interfaces is a much larger problem than is currently soluble with XML/XUL (e.g. see [14]).

References

1. Walker, A., Brewster, S. A., McGookin, D., Ng, A.: Diary in the Sky: A Spatial Audio Display for a Mobile Calendar. In: Alan Blandford, J. Vanderdonckt, and P. Gray, (eds.): People and Computers XV-Interaction without Frontiers: Proceedings of IHM/HCI 2001. Springer-Verlag (2001) 531-539
2. Dunlop, M. D., Brewster, S. A.: The Challenge of Mobile Devices for Human Computer Interaction. Personal and Ubiquitous Computing 6 (2002) 235-236
3. Forrester Research. (2004) Statistics for Mobile Commerce.[Online]. Available: http://www.epaynews.com/statistics/mcommstats.html#36
4. Chittaro, L., Dal Cin, P.: Evaluating Interface Design Choices in WAP Phones: Navigation and Selection. Personal and Ubiquitous Computing 6 (2002) 237-244
5. Juhn, J. H. (2004) Value chains for business-to-business m-commerce.[Online]. Available: http://www.tuta.hut.fi/studies/Courses_and_schedules/Isib/TU-91.167/Old_seminar_papers/Old_seminar_papers.php
6. NASA Human Performance Research Group: Task Load Index (NASA TLX) v1.0 computerised version. NASA Ames Research Centre (1987)
7. Hart, S., Staveland, L.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: P. Hancock and N. Meshkati, (eds.): Human Mental Workload. North Holland B.V (1988) 139-183
8. Card, S. K., Moran, T. P., Newell, A.: The Psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, Hillsdale, NJ (1983)
9. Dix, A. J., Finlay, J. E., Abaowd, G. D., Beale, R.: Human-Computer Interaction. Prentice Hall (1998)
10. Brewster, S. A.: Overcoming the Lack of Screen Space on Mobile Computers. Glasgow University, Department of Computing Science Technical Report (2001)
11. Vickers, P., Alty, J. L.: Using Music to Communicate Computing Information. Interacting with Computers 14 (2002) 435-456
12. Vickers, P., Alty, J. L.: When Bugs Sing. Interacting with Computers 14 (2002) 793-819
13. Chittaro, L., Dal Cin, P.: Evaluating Interface Design Choices on WAP Phones: Single-choice List Selection and Navigation Among Cards. In: Proc. Proc Mobile HCI 2001: Third International Workshop on Human Computer Interaction with Mobile Devices (2001)
14. Thevenin, D., Coutaz, J.: Plasticity of User Interfaces: Framework and Research Agenda. In: Proc. INTERACT '99 - Seventh IFIP Conference on Human-Computer Interaction (1999) 110-117

Personalization-Based Optimization of Web Interfaces for Mobile Devices

Michael Hinz, Zoltán Fiala, and Frank Wehner

Dresden University of Technology
Heinz-Nixdorf Endowed Chair for Multimedia Technology
Mommsenstr. 24
D-01062 Dresden
{michael.hinz, zoltan.fiala, frank.wehner}@inf.tu-dresden.de

Abstract. Developing personalized applications for the ubiquitous Web assumes to provide different user interfaces addressing heterogeneous capabilities of device classes. Major problems are the lack of sufficient presentation space and the diversity of interaction techniques, both requiring adaptive intelligent user interfaces. To meet this challenge this paper introduces an approach for the personalization-based optimization of Web interfaces for mobile devices. On the basis of a user model different adaptation issues are discussed. Firstly, static adaptation mechanisms affecting the structure of Web documents as well as layout managers enabling a device independent definition of Web presentations for heterogeneous devices are introduced. Then an interactive mechanism for dynamically predicting user preferences for hiding unnecessary information through content adaptation is presented. As a proof of concept an architecture realized by a pipeline-based document generator was developed for static/dynamic adaptation, which is partly explained in this paper.

1 Introduction

Providing personalized information becomes a significant challenge of today's Web development. The raising number of users with an increasing variety of mobile devices requires the creation and publication of content customized for different user preferences and platforms. A major problem is the diversity of display capabilities and interaction techniques provided by mobile clients, which establishes the need for adaptive intelligent user interfaces that automatically adjust their content to those heterogeneous requirements. However, existing document formats (such as HTML, cHTML or WML) are hardly suitable for engineering personalized ubiquitous Web applications, as they do not provide mechanisms for describing the adaptive behavior of content pieces in a generic way.

Existing approaches for displaying Web content on mobile devices mostly focus on restructuring or clipping existing pages according to static guidelines [1], [2], [4]. However, including the user's changing interests in this process enables not only a better personalization but also an optimized utilization of the available presentation area.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 204–215, 2004.

The paper is structured as follows. After addressing related work in Section 2 a short overview of our component-based document model for personalized ubiquitous Web presentations is given. Section 4 deals with different aspects of adaptation supported by the document format, and gives a short introduction to the user model. On this basis static adaptation in dependency of user and device properties, dynamic adaptation in dependency of user preferences and an automatic layout adjustment mechanism are discussed. The implemented system architecture is explained in Section 5. Section 6 concludes the paper and suggests future research directions.

2 Related Work

Recently, different solutions for adapting Web presentations and applications to mobile devices have emerged. Basically two main approaches can be distinguished. The first one adjusts existing Web pages (mostly HTML) to the limited display and interaction capabilities offered by mobile devices. The second one aims at building personalized ubiquitous Web applications “from scratch” and considers device (and user) adaptation already during the specification and implementation process.

Different mechanisms for automatically adjusting existing desktop Web pages to mobile browsers have been developed. Some solutions, e.g. Microsoft’s Pocket Internet Explorer [1] or Opera for Smartphones/PDAs [2] resize large Web pages to fit into the small displays of mobile clients. Even though all information from the original page is displayed, it is reformatted in order to eliminate horizontal scrolling. The disadvantage of this approach is a presentation often featured with unnecessary information or layout fragments. Therefore, Web clipping techniques have emerged which firstly analyze the structure of Web pages. By discovering priorities, page fragments are classified as either important or unimportant, and the latter are excluded from the “clipped” presentation. Two strategies for defining priorities exist. The first one uses intelligent algorithms to automatically classify page fragments [3], [4]. The second strategy [5], [6] requires a manual definition of priorities. As further interesting approaches we mention HANd [7] and SmartView [8] which structure the original Web page into zones. Through automatically generated summary pages or thumbnails every zone can be reached via navigation. The advantage of those techniques is that no information is clipped since by navigation every zone can be reached. Still, extra navigation is required and by splitting a page the overview gets lost. Therefore the user’s mental load rises. A similar approach for text browsing [9] enables the summarization of texts with an "accordion" display technique.

The main advantage of the approaches mentioned above is that they are principally suitable for adapting arbitrary Web pages. However, evaluations ([10]) show that it is often impossible to predict (or enforce) the result of the transformation process and that in many cases erroneous output pages are provided. Furthermore, since all these approaches operate on the HTML-based presentation view of their input pages, adaptation is restricted to the exclusion or rearrangement of content pieces. On the other hand, we claim that effective device adaptation has to be already considered during the conceptual and navigational design of Web applications.

Recently, different approaches for modeling and engineering ubiquitous personalized Web systems have emerged. Among the most significant ones we mention WebML [11] and Hera [12]. However, all these approaches focus on the conceptual modeling

and design of hypermedia applications, not supporting the flexible reuse of adaptable implementation artifacts. Furthermore, device adaptation is not a central aspect of these approaches. To fill this gap, the project AMACONT [13] recently introduced a component-based document format for personalized ubiquitous Web presentations [14]. It focuses not on the conceptual design of Web applications, but on the challenge to reuse adaptable implementation artifacts. In this paper a detailed overview of personalization issues (with a special focus on device adaptation) is given.

3 The Document Model

In the Amacont approach Web sites are composed of configurable Web components [14]. These components are instances of an XML grammar representing adaptable content on different abstraction levels. Web sites are constructed by aggregating and linking components to complex document structures. During Web page generation these abstract document structures are translated into Web pages in a concrete output format, adapted to a specific user model or client device, respectively.

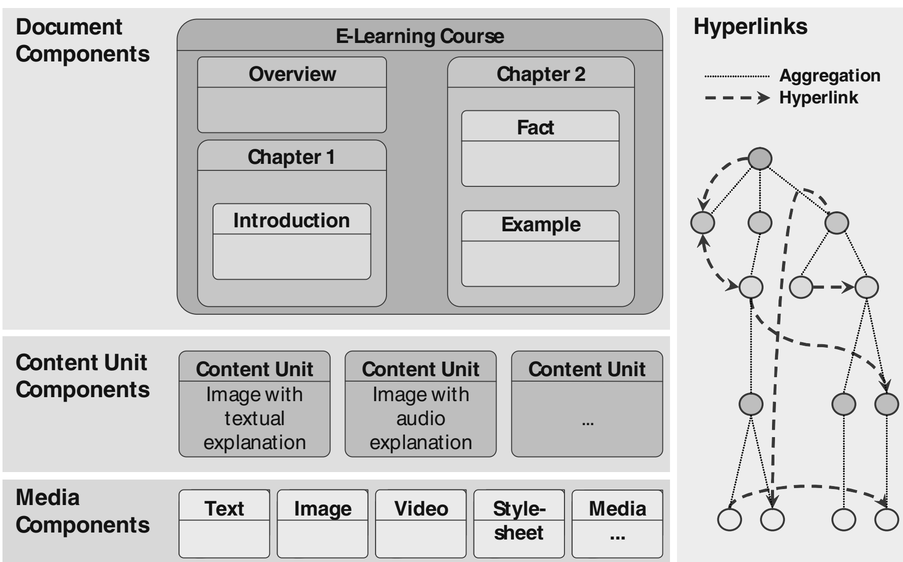


Fig. 1. The document model

The lowest abstraction level introduces *media components* that encapsulate concrete media assets. These comprise text, structured text (e.g. HTML), images, sound, video, Java applets and may be extended arbitrarily. Besides MPEG7-based technical properties additional content management information is provided, too.

On the second level media components belonging together semantically - e.g. an image with textual description - are combined to so called *content unit components*. Defining such collections is a key factor of reuse. The spatial adjustment of contained

media components is described by client-independent layout properties abstracting from the exact resolution and presentation style of the current display (Section 4.3).

Thirdly, *document components* are specified as parts of Web presentations playing a well defined semantic role (e.g. a news column, a product presentation or even a Web site). They can either reference content units, or aggregate other document components. The resulting hierarchy describing the logical structure of a Web site is strongly dependent from the application context. Again, the spatial adjustment of subcomponents is described in a client-independent way.

Finally, the orthogonal hyperlink view defines links spanned over all component levels. Uni- and bidirectional typed hyperlinks based on the standards XLink, XPath and XPointer are supported. For a detailed introduction to the document model the reader is referred to [14].

4 Adaptation Support

The component-based document format aims at supporting adaptation by two mechanisms [15]. Firstly, it enables to encapsulate adaptation logic in components on different abstraction levels. Secondly, it allows describing the visual aspects of components by client-independent layout descriptors that can be automatically adapted to different output formats. Both adaptation aspects can be declared by attaching specific adaptation metadata to components. During document generation, this metadata is evaluated according to an XML-based user model and the corresponding adaptation processes are performed.

Furthermore, two types of adaptation or personalization can be distinguished: adaptability and adaptivity. Adaptability (also known as static adaptation) means that the generation process is based on available information that describes the situation in which the user will use the generated presentation [16]. Adaptivity (also mentioned as dynamic adaptation) is the kind of adaptation included in the generated adaptive hypermedia presentation. To put it simple, in the second case the hypermedia presentations themselves change while being browsed. This dynamic nature of adaptivity is supported by feedback mechanisms updating the user model according to the user's interactions with the presentation.

This section provides an overview of AMACONT's versatile adaptation capabilities. Firstly, the structure of the user model is depicted which is used across all examples. Then, different aspects of static and dynamic personalization are described in detail. All introduced adaptation examples aim at optimizing Web presentations to mobile end devices.

4.1 The User Model

The adaptation of components happens according to an XML-based user model. This is composed of a number of profiles that can be seen in Fig. 2 Each profile relies on CC/PP (Composite Capability / Preference Profiles), an RDF grammar for describing device capabilities and user preferences in a standardized way [17]. However, as being a general grammar, CC/PP makes no assumptions on concrete resource

characteristics. Therefore, an XML-based schema was developed for each profile. By adding new profiles the user model can be extended arbitrarily.

The first part (*IdentificationProfile*) of the user model contains information to identify users. Besides a set of general properties (name, email etc.), arbitrary extensions are allowed. Technical properties and capabilities of users' client devices are stored in *DeviceProfile*. It is represented on the basis of the WAP User Agent Profile (UAProf [18]) providing a common vocabulary for WAP devices. To support also other mobile devices (e.g. PDAs), specific extensions of UAProf have been made. Furthermore, as usually there are much more users than devices, it is also possible to reference separately stored device profiles.

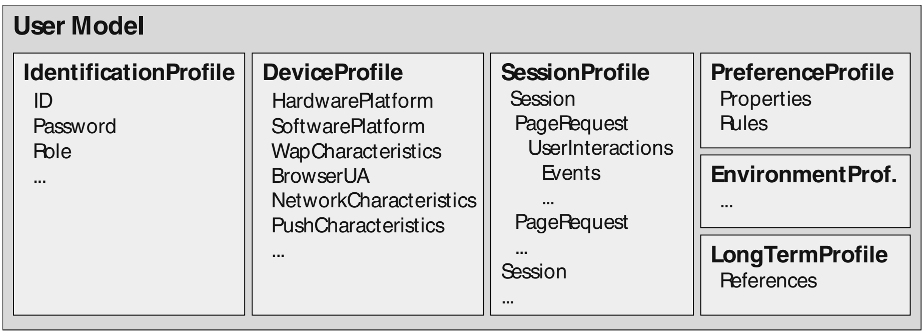


Fig. 2. The user model

The *SessionProfile* integrates user interactions by grouping them to page requests and sessions. It stores past user interactions in the form of events related to data acquisition objects (see section 4.4). Based on this interaction history list the user modeling process generates new knowledge about the user in term of rules (see Section 4.4). Those rules are stored in the *PreferenceProfile* and used by the document generator to adapt the content of a web page to user preferences. The last two profiles are placeholders for upcoming research. *EnvironmentProfile* will provide information about the context and location of the user for supporting location based services. *LongtermProfile* will have a bridging function between a special user model and comprehensive models containing information about all users of the system. E.g. the user class membership of a user will be represented by this profile in order to reduce server load by handling groups of users together.

4.2 Static Adaptation in Dependency of User and Device Properties

The document format described in Section 3 supports personalization by encapsulating adaptive behavior in components on different abstraction levels. Firstly, adaptation is required on the level of media components in order to consider various client capabilities or other technical preferences (e.g. bandwidth, color depth, etc.) by providing alternative media instances with varying quality. Secondly, on the level of content units the number, type and arrangement of inserted media components can be adjusted. Consider the case of two online-shop customers, one of them preferring detailed textual descriptions, the other visual information. The presentation for the

first user might include content units containing text objects, for the other one rather images or videos. Thirdly, personalization of document components concerns the adaptation of the whole component hierarchy, which results in different subcomponent trees for different user preferences and/or device capabilities. Finally, adapting hyperlinks enables personalized navigation structures within the generated Web presentation.

In order to describe adaptive behavior in a generic way, each component may include a number of variants. As an example, the definition of an image component might include two variations for color and monochrome displays. Similarly, the number, structure, arrangement and linking of subcomponents within a document component can also vary depending on device capabilities or user properties. The decision, which alternative is selected, is made during document generation by an XSLT stylesheet according to a certain selection method which is described in the component's header. Such selection methods are chosen by component developers at authoring time and can represent arbitrary complex conditional expressions parameterized by user model parameters. This separation of describing variants (in the component body) and adaptation logic (in the component header) allows reusing a given component in different adaptation scenarios. The XML code below demonstrates the definition of a document component's variants and a selection method. In a Web presentation offering video tapes, different content depending on the bandwidth of the user's device is presented.

Table 1. Defining component variants (left) and selection methods (right)

```
<AmaDocumentComponent name="Film">
  <MetaInformation>
      ...
  </MetaInformation>
  <Variants>
      <Variant name="Video_Trailer">
          ...
      </Variant>
      <Variant name="Cover_Picture">
          ...
      </Variant>
  </Variants>
</AmaDocumentComponent>
```

```
<AdaptiveProperties>
  <If>
 <Expr operator=”greaterThan”>
      <UserModelParam>
            Bandwidth
      </UserModelParam>
      <Const>64000</Const>
 </Expr>
      <Then res="Video_Trailer"/>
      <Else  res="Cover_Picture"/>
 </If>
<AdaptiveProperties>
```

The processing XSLT style sheet substitutes the integer variable "Bandwidth" by its value from the current user model, performs the selection method and determines the proper variant of the "Film" component. As this variant might also have varying subcomponents, the style sheet works recursively. The XML-grammar for selection methods allows the declaration of user model parameters, constants, variables and operators, as well as complex conditional expressions of arbitrary depth. The processing XSLT stylesheets act as an interpreter for this "selection method language".

4.3 Automatic Layout Adaptation

In order to describe the presentation of component-based Web documents, AMACONT allows attaching XML-based layout descriptions to components.

Inspired by the layout manager mechanism of the Java language (AWT and Swing) and the abstract user interface representations of UIML [19] and XIML [20], they describe a client-independent layout allow abstracting from the exact resolution of the display or the browser's window. Note that layout managers of a given component only describe the presentation of its immediate subcomponents, which encapsulate their own layout information in a component-based way.

At current time four layout managers can be defined. *BoxLayout* allows multiple components to be laid out either vertically or horizontally. *BorderLayout* arranges components to fit in five regions: north, south, east, west, and center. *GridTableLayout* enables to lay out components in a grid with a configurable number of columns and rows. Finally, *OverlayLayout* allows to present components on top of each other.

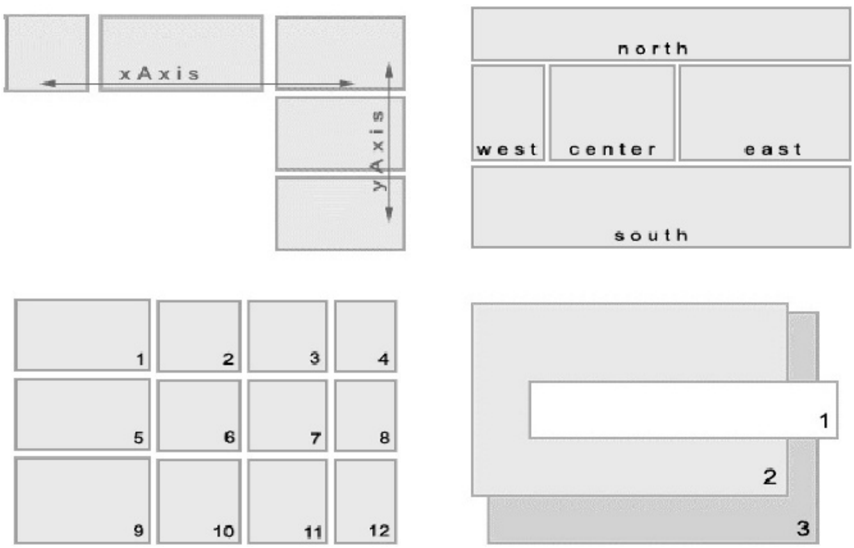


Fig. 3. Layout managers: upper left: BoxLayout, upper right BorderLayout, lower left: GridTableLayout, lower right OverlayLayout

Layout managers are formalized as XML elements with specific attributes. Two kinds of attributes exist: *layout attributes* and *subcomponent attributes*. Layout attributes declare properties concerning the overall layout and are defined in the corresponding layout tags. As an example the *axis* attribute of BoxLayout determines whether it is laid out horizontally or vertically. On the other hand, subcomponent attributes describe how each referenced subcomponent has to be arranged in its surrounding layout. Table 2 summarizes the possible attributes of BoxLayout by describing their names, role, usage (required or optional) and possible values.

The optional attribute *wml_visible* determines whether in a WML presentation the given subcomponent should be shown on the same card. If not, it is put onto a separate card that is accessible by an automatically generated hyperlink, the anchor text of which is defined in *wml_description*. This mechanism of content separation and navigation adaptation is used since the displays of WAP capable mobile phones are very small.

Table 2. Example: layout attributes of the BoxLayout manager

Layout Attributes	Meaning	Usage	Values
axis	orientation of the BoxLayout	req.	xAxis | yAxis
space	space between subcomponents	opt.	percent or absolute
width	width of the whole layout	opt.	percent or absolute
height	height of the whole layout	opt.	percent or absolute
border	width of border between subcomp.	opt.	percent or absolute
Subcomponent Attributes			
align	horizontal alignment of subcomp.	opt.	left | center | right
valign	vertical alignment of subcomponent	opt.	top | center | bottom
ratio	space taken by subcomponent	opt.	percent
wml_visible	show on same WML card?	opt.	boolean
wml_desc	link description for WML	opt.	string

The exact rendering of media objects happens during document generation time by XSLT stylesheets that transform components with such abstract layout properties to specific output formats. Three stylesheets for converting those descriptions to XHTML, cHTML and WML output have been realized.

4.4 Dynamic Adaptation Issues

The mechanisms described above support adaptability by adjusting Web presentations to (mostly) static user and device properties. However, in order to realize dynamic adaptation (or adaptivity), they have to be extended by additional feedback mechanisms. User interactions have to be captured on the client and sent back to the server in order to update the user's preference profile, i.e. to automatically generate adaptation rules according to the user's browsing behavior. In contrast to other approaches (e.g. [3], [5], [6]), this allows to adjust Web presentations to even dynamically changing user interests.

Note that this strategy can be effectively used for optimizing Web pages on mobile devices with limited presentation space. As an example, take the case of an interactive multimedia Web presentation allowing to perform interactions on selected media items. A user being more interested in textual information (due to the limited display capabilities of his browser) could collapse images and enlarge texts. A corresponding learning algorithm could recognize this and generate the appropriate adaptation rules which automatically collapse all images for the user's display.

A further possibility is to provide observed media components with a special semantic meaning in order to predict semantic user preferences. Let us take the case of an online product presentation where a user enlarges a picture containing technical features of a selected product and then changes to the next product. The system could establish a rule that the user is interested in technical details and generate the next product presentation according to this rule.

Acquire Interactions

In order to observe users' browsing behavior, our developed system allows to track interactions that are performed on media components included in a Web page. During

server side document generation specific code fragments (implemented as JavaScript or JScript functions) are embedded and configured for each media component to be observed. They allow capturing user interactions on the client side and sending them back to the server, where they are stored in history lists (session profile). Acquirable interactions are listed in Table 3.

Table 3. Acquirable interactions of observed components

observed component	acquirable interactions
video and audio component	started, paused at, stopped at
image component	minimized, maximized, printed
scroll text component	scrolling time, end reached
toggle text component	enlarged, collapsed
pop up text component	pop up

In order to make media components observable, component authors have to provide them with specific metadata. Hence, semantic metadata in the form of attribute-value pairs (e.g. content="technical details") can be attached to them. Thus, the semantic preferences of user's interacting with those objects can be predicted.

Processing Interactions

By evaluating interactions, suggestions on users' preferences and knowledge can be made and parts of the user model can be updated or specialized. In our developed prototype application focusing on product presentation this specialization is performed by the incremental learning algorithm CDL4 (Complementary Discrimination Learning [21]). The algorithm was approved as very useful in adaptive multimedia product presentations in an earlier project of the authors' research group [22].

CDL4 utilizes decision lists in order to describe user models. A decision list is a series of simple rules describing user preferences. As an example, the following decision list claims that the user is not interested in multimedia information about actors other than the main actor:

$$[((actor \neq mainActor)\ \Lambda\ (medium \neq text) \rightarrow noInterest),$$
$$(default \rightarrow interest)]$$

If no rules from earlier sessions exist, CDL4 starts with a minimal default decision list (see second line in the example above) in the beginning of each user session. According to the user's interaction behavior, this is extended (specialized) in an incremental way.

Interactions stored in the session profile are transformed to so called training instances. Training instances are also formed as single decision rules and serve as the input for the CDL4 algorithm. For instance, if the user enlarges a picture component containing the biography of a supporting movie actor, the server generates following training instance:

$$[biography, supportingActor, picture \rightarrow interest]$$

Each time a new training instance is provided, the algorithm has to check whether its current decision list already covers this new instance. If yes, the decision list remains unchanged. Otherwise, the algorithm learns this new instance and updates (specializes) the corresponding decision list by changing an existing rule or inserting a new one. In our example, the update decision list would look like this:

$$[((actor \neq mainActor)\ \Lambda\ (medium \neq text)\ \Lambda\ (medium \neq picture) \rightarrow noInterest),\ (default \rightarrow interest)]$$

At the user's next document request, the inserted media components are configured according to the new rules. For more details on CDL4 the reader is referred to [21].

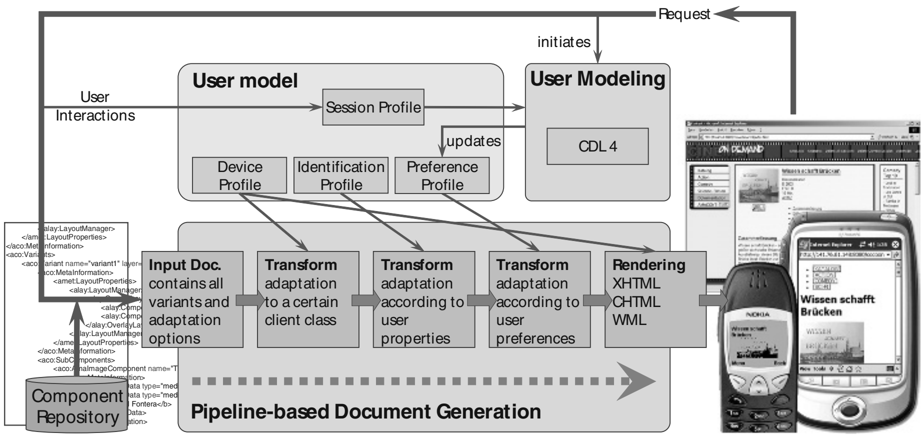


Fig. 4. Pipeline-based document generation

5 Generating Adaptive Web Documents

Document generation aims at transforming complex component structures to Web pages adapted to user properties and preferences as well as device profiles. It is performed in a stepwise, pipeline oriented way (Fig. 4). For each user request, a complex document encapsulating all possibilities concerning its content, layout, and structure is retrieved from a component repository. According to the user model (containing also the device profile), it is subdued to a series of XSLT transforms, each considering a certain adaptation aspect by the configuration and selection of component variants (see Section 4.2).

Fig. 4 shows a possible scenario with three steps, namely adaptation to a certain client class (e.g. PDA, cell phone or notebook), then to static user properties (age, gender, knowledge level, etc.) and finally to semantic user preferences (e.g. interests, media preferences).

In this scenario the first two adaptation steps are performed according to the variant selection mechanism described in Section 4.2. Thus, the hierarchy of components is adjusted to static user properties and device profiles.

The third step, namely dynamic adaptation according to changing user preferences affects not the aggregation hierarchy of the overall component structure but the presentation parameters of single media components. For example, an image can be inserted minimized or maximized, a text can be presented in a short or in a long form, or even videos can be started automatically. These decisions are made by the CDL4-algorithm according to the rules stored in the preference profile.

After the component hierarchy to be presented and the parameters of media objects have been determined, the resulting adapted document has to be transformed to a specific output format (XHTML, cHTML, WML etc.). According to the layout managers described in Section 4.3, this rendering happens automatically. Moreover the data acquisition objects for tracking user interactions are included in this transformation step, too. Again, they enable to track user interactions in the newly generated presentation. This loop enables a dynamically adaptation process with an always up-to-date user model.

6 Conclusion and Future Work

In this paper an overview of the adaptation issues provided by the XML-based document model and the system architecture of the AMACONT project was given. Both static adaptation issues based on user and device properties and dynamic personalization aspects according to dynamically changing user preferences were discussed. Furthermore, a pipeline-based document generator was introduced for performing those adaptations in a stepwise way. We have shown how the Web interface of mobile devices can be optimized by those personalization techniques. Especially the observation of users and the prediction of their preferences enabled an automatic prioritization of content and therefore the hiding of unnecessary information from the user.

Future work concentrates on the authoring process of dynamically personalized Web documents for heterogeneous mobile devices. A modular framework for creating and configuring components in different stages of the authoring process is being built. Furthermore, performance aspects of the system architecture will be addressed, too. Since dynamic adaptation mechanisms cause significant server load, optimizing the performance seems to be an important effort when handling lots of users. Initial tests showed that the number of requests and the structure of existing rules play an important role when the system manages dynamic adaptation. Reducing rules representing user preferences to a minimum could improve the overall performance.

References

1. Pocket IE: http://www.microsoft.com/windowsmobile/products/pocketpc/default.mspx
2. Opera's Small Screen Rendering: http://www.opera.com/products/smartphone/smallscreen
3. Gomes, P.; Tostao, S.; Goncalives, D.; Jorge, J.: Web Clipping: Compression Heuristics for Displaying Text on a PDA. In Proceedings of the Mobile HCI'01, 2001.
4. Bickmore, T.; Schilit, B.: Digestor: Device-independent Access to the WWW. In Proceedings of the WWW6 conference, 1997.

5. Hori, M.; Kondoh, G.; Ono, K.; Hirose, S.; Singhai, S.: Annotation-based Web Content Transcoding. In Proceedings of the WWW9 conference, Amsterdam (Netherlands), 2000.
6. Palm Web Clipping: http://www.palmos.com/dev/tech/webclipping/resources.html
7. González-Castano, F.; Anido Riffón, L.; Costa Montene-gro, E.: A New Transcoding Technique for PDA Browsers, Based on Content Hierarchy. In Proceedings of the 4th International Symposium on Mobile HCI, 2002.
8. Milic-Frayling, N.; Sommerer, R.: SmartView: Flexible viewing of web page contents. Poster presentation at the WWW11 conference, 2002.
9. Buyukkokten, O.; Garcia-Molina, H.; Paepcke, A.: Seeing the whole in parts: Text summarization for web browsing on handheld devices. In Proceedings of the WWW10 conference, Hong-Kong, 2001.
10. Chen, Y.; Ma, W.; Zhang, H.: Detecting Web Page Structure for Adaptive Viewing on Small Form Factor Devices. In Proceedings of the WWW12 conference, 2003.
11. Ceri, S.; Fraternali, P.; Bongio, A.; Brambilla, M.; Comai, S.; Matera, M.: Designing Data-Intensive Web Applications. Morgan Kaufmann, ISBN 1558608435, 2003.
12. Vdovjak, R.; Frasincar, F.; Houben, G.J.; Barna, P.: Engineering Semantic Web Information Systems in Hera. In Journal of Web Engineering, Vol.2 No.1&2, RP, 2003.
13. AMACONT Project: http://www-mmt.inf.tu-dresden.de/english/Projekte/AMACONT/
14. Fiala, Z.; Hinz, M.; Meißner, K.; Wehner, F.: A Component-based Approach for Adaptive, Dynamic Web Documents. In Journal of Web Engineering, Vol.2 No.1&2, RP, 2003.
15. Fiala, Z.; Hinz, M.; Wehner, F.: An XML-based component architecture for personalized adaptive web applications. In Workshop Personalisierung mittels XML-Technologien, Berliner XML Tage, Pages 370 - 378, 2003.
16. Fiala, Z.; Hinz, M.; Houben, G.; Frasincar, F.: Design and Implementation of Component-based Adaptive Web Presentations. In Proceedings of the 19th Annual ACM Symposium on Applied Computing, 2004.
17. Klyne, G.; Reynolds, F.; Woodrow, C.; Ohto, H.; Hjelm, J.; Butler, M.; Tran, L.: Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0. W3C Recommendation 15 January 2004.
http://www.w3.org/TR/2004/REC-CCPP-struct-vocab-20040115/
18. Wireless Application Group: User Agent Profile Specification. Open Mobile Alliance WAP Forum 2001.
19. User Interface Markup Language (UIML): Specification Draft Language, Version 3.0, February 2002. http://www.uiml.org/specs/
20. Puerta, A; Eisenstein, J.: XIML: A Universal Language for User Interfaces. In Proceedings of the Conference on Intelligent User Interfaces, 2002.
21. Shen, W.: An efficient Algorithm for Incremental Learning of Decision Lists. Technical Report, USC-ISI-96-012, University of Southern California, 1996.
22. Jörding, T., Meissner, K.: Intelligent Multimedia Presentations in the Web: Fun without Annoyance. In Proceedings of the WWW7 conference, Brisbane, 1998.

Mixed-Initiative, Trans-modal Interface Migration

Renata Bandelloni, Silvia Berti, and Fabio Paternò

I.S.T.I-C.N.R. via G.Moruzzi, 1,
56124 Pisa, Italy
{Renata.Bandelloni, Silvia.Berti, Fabio.Paterno}@isti.cnr.it

Abstract. This paper presents our solution to supporting runtime migration of Web application interfaces among devices offering different interaction modalities, in particular graphic to vocal platform migration and vice versa. Migrating between platforms implies keeping track of the user interactions in order to retrieve the runtime state of the interface and maintaining interaction continuity on the target device. The system can serve user-issued migration requests containing the identifier of the selected target device, and can also automatically start the migration procedure when environment conditions require it. In automatic migration the target platform has to be automatically selected as well. To this aim, we consider devices belonging to a restricted environment and have defined selection rules in order to identify the most suitable available target for the ongoing migration.

1 Introduction

The wide availability of new mobile devices supporting Internet access, offering various interaction capabilities, raises the need for applications able to support different interaction modalities. On any given day, users are surrounded by many different interactive platforms at work, at home, and even while walking along the road. User mobility accompanied by various devices raises the need for some sort of application interface mobility that allows users to change the device they are interacting with while moving from one environment to another, or just because the resources (such as the battery) of the current mobile device have been depleted. Such scenarios raise the need for multi-platform migration services that are able to follow users through the changing contexts by transferring amongst different devices at run time.

We analyzed the potentialities of our model-based approach to perform an adaptive interface migration [1], addressing devices with different features, able to support graphic navigation of Web sites. In this work we address two novel issues for this approach:

- introduction of migration with modality change, thus allowing users to migrate from a graphical interface to a vocal interface or vice versa;

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 216–227, 2004.

- support for migration that can be activated not only on user request but also by the automatic detection that a system is no longer able to support the user (for example, because the battery has expired or the system is no longer connected).

The solution to these issues is important because allows users interacting with a typical graphical Web browser, to continue the interaction through a different device and a different interaction modality. This is the case of a user navigating the Web through a PDA or Desktop PC and migrating the application to a mobile phone supporting only vocal interaction. Apart from migration on user demand, we also introduce the novelty of automatic migration, which implies automatic recognition of nearby devices associated to the same user as well as the ambient conditions leading to migration.

Graphic and vocal interactions rely on different interaction techniques because of the differences between the associated media. In graphic browsers, many tasks can be supported concurrently, all at once in a page, and the user can freely decide which one to perform. Vocal navigation imposes a serialisation of dialogues. At any time only one interaction is available, even if users can choose to move at different points of the dialogue structure. Such differences imply a different way to structure the concrete interface and choose the interface elements.

Our migration service applies to Web applications whose interface has been developed through the model-based approach supported by the TERESA tool [5]. This provides semantic information associated with the user interface implementation that can be exploited at run-time to support migration in such a way as to maintain interaction continuity and consider usability design criteria.

In section 2 we discuss related work. Next, we introduce a couple of scenarios highlighting the issues that we aim to address. Then, we discuss the solution developed to obtain migratory interfaces through underlying transformations and processing. This is followed by the description of the architecture of the migration service highlighting how it can support both user-activated and system-activated migration. We also provide more detail showing how the trans-modality migration is achieved along with an example of application. Some concluding remarks and indications for future work conclude the paper.

2 Related Work

Run-time adaptation of user interfaces to different device capabilities raises many issues. A framework describing such issues is provided in [2]. PUC [6] is an environment that supports the downloading of logical descriptions of appliances and the automatic generation of the corresponding user interfaces. The logical description is performed through templates associated with design conventions, which are typical design solutions for domain-specific applications. The application area of this approach is limited to the home domain where devices require similar interfaces. Aura [3] is a project whose goal is to provide an infrastructure that configures itself automatically for the mobile user. When a user moves to a different platform, Aura

attempts to reconfigure the computing infrastructure so that the user can continue working on tasks started elsewhere. In this approach, suppliers provide the abstract services, which are implemented by just wrapping existing applications and services to conform to Aura APIs. For instance, Emacs, Word and NotePad can each be wrapped to become a supplier of text editing services. So, the different context is supported through a different application for the same goal (for example, text editing can be supported through MS Word or Emacs depending on the resources of the device at hand). Our work follows a different approach where the application is still the same but the interactive part is adapted to the new device.

A taxonomy of task for voice interfaces to Web pages [7] has been proposed with the aim to obtain a voice navigation that is not a mere substitution of graphic navigation, but is tailored on specific voice interaction features. The paper focalizes on VoiceXML interface design, in our work we take into consideration both vocal and graphic navigation and make a comparison between them in order to allow a user to change interaction modality while changing device.

An example of transformation of a Web site developed in HTML into a VoiceXML application is presented in [4]. In this paper the original Web site is analyzed and redesigned in a dialog model style, by means of a finite state diagram, then the model is implemented in VoiceXML. The model of the vocal version of the original site is manually built; authors are working on the automatic remodelling of the Web site, basing on a syntactic and semantic HTML files analysis. In our approach we consider interfaces generated from task models. This semantic information is exploited also in the migration service in order to identify what part of the target interface to activate and to associate it with the state of the user interactions performed so far.

Perez-Quinones et al. [7] describe a multimodal interface architecture that allows for combining speech, pen and touch-tone digit interaction in noisy mobile environments. The proposed system allows users to interact with an application using more than one modality at once. The system was evaluated through an example application. One result is the confirmation that some kinds of tasks are more appropriate for a specific input modality. In that work, the user can access different interaction modalities at the same time, over a single device.

3 Scenarios

In this section we present two scenarios to underline the features of the multi-modal migration service. The first one concerns a restaurant booking application and is an example of graphic to vocal migration.

Friday Morning, Louis is at home and wants to organize a dinner with his friends for the evening. He turns on his personal computer and opens the Official Web site of the town. He accesses the restaurant main page, from which he starts selecting restaurants one by one, in order to check the menu of the day. While Louis is selecting the Mermaid restaurant main page, he realizes that it is getting late and has to leave and go to work, hence requires the migration to his mobile phone. Louis can now turn off the computer keep interacting with the application in vocal mode. The vocal interface

remembers the selected restaurant to Louis and tells him the different options he can go through. Louis asks to hear the menu of the day, then he asks to go back to the main restaurant options and asks for booking a table. The system asks Louis to say his name, selects the preferred menu, and specifies the date and time for booking. Finally, the system repeats all the information inserted to Louis, asking confirmation. Louis confirms the booking and keeps walking to his office, enjoying the thought of the dinner with his dearest friends.

The second scenario concerns a typical agenda application and is an example of vocal to graphic migration. Monday morning, George is driving in his car when an accident blocks the road. George will be late for work, so he decides to access the Agenda Application through the car voice system to check his schedule for the day. The voice synthesizer welcomes George to the Agenda Application and tells him all the available operations. George says "Today's *schedule*" to check the appointments fixed for the day and under a further system request, he specifies that he wants to hear the appointments scheduled for the morning. The synthesizer says that he has two appointments scheduled in the morning and the first one is a 10:00 meeting with the project coordinator. George asks for more details, meanwhile he arrives at work. As soon as he turns off the car, the application migrates automatically from the voice car system to the PDA that George has in his pocket. George starts running to his office to collect important documents for the meeting and use his PDA to check in which room the meeting is to be held. In the above scenario, the vocal interaction is supported by a voice car system. In our analysis, we take into consideration a vocal interface accessed through a mobile phone. Diverse voice car kits connect to the car owner mobile phone, as soon as the vehicle is turned on allowing automatic call answering. With such kind of equipment, migration can take place from the PDA to the mobile phone and vice versa, giving the user the feeling that only the phone and the car are involved. In particular, the user will not hear the phone ring, announcing graphic to vocal migration, because of the automatic call answer feature, and will be able to continue interacting, without any supplementary action to receive the call.

4 Migration Service Approach

The interface migration is obtained through different interface versions (one for each platform). When the interface migrates then the migration service is able to activate the version for the target device at the point where the user left the source device and maintain the state resulting from the previous interactions in the new device.

Our migration service applies to Web applications whose interfaces have been developed through the model-based approach integrated in the tool TERESA. The interface generation through the TERESA approach, starts with the development of the nomadic task model that describes the application interface in terms of user activities. Platform specific task models are obtained analysing the nomadic one, extracting the tasks supported by the specific platform. Each refined task model is used to generate the Abstract User Interface (AUI), where the interface is described in terms of presentations. Each presentation contains: a set of Interactors, giving an

abstract description of the objects that will be used to implement corresponding tasks, and composition operators, providing declarative indications on how to compose interactors (grouping, hierarchy, relation, …). Each presentation is associated with the set of tasks that it supports. The last step is the generation of the final user interface according to design criteria that take into account the platform selected. It can be generated in XHTML, XHTML Mobile Profile, Java, or VoiceXML. In this paper we are considering XHTML and VoiceXML languages, in order to address the trans-modality migration.

The logical descriptions obtained through the interface generation process are used by the Migration Server in order to compare source and target platform version of the interface, to identify the presentation for the target platform and keeping user interaction continuity, when activating the interface on the target device. When a migration is required by the user, or automatically triggered, the server retrieves the last URL loaded on the source device, hence extracts the presentation describing the migrating page, from the AUI of the corresponding interface. At this point, the server accesses the AUI describing the interface for the target platform type.

The server uses the AUIs, to search for the target presentation that is the most similar to the source one. Similarity is calculated in terms of supported tasks: the higher number of tasks the source and target presentation share, the more similar the presentations are. This similarity criterion can lead to ambiguity in case more than one target presentations share the same number of tasks with the source one, having the same similarity degree. The conflict is solved identifying the target presentation supporting the task associated with the interaction object last modified by the user on the source device, since the user is most likely to continue interaction from that point.

Once the target presentation has been identified, the target page is immediately identified, since a one to one mapping exists between presentations and pages. Next step is to calculate the state of the objects contained in the target page, in order to keep interaction continuity. In this phase, we consider objects implementing corresponding tasks in the source and target page. In different versions of the interface obtained for different platform types, the same task can be implemented by means of different interaction objects. In particular, while comparing graphic and vocal platforms, we have VoiceXML objects in one version and HTML objects in the other one. For example, the graphical interface task that performs a selection action can be implemented by radio button while, in the vocal interface, this can be obtained through DTMF (Dual Tone Multi Frequency) menu voice, and the selection can be performed through keypads. Another interesting example is given when in the logical description of the presentations there are two or more *control* interactors that enable the access to other application pages: in the graphical interface they can be implemented by buttons or links while in the voice interface they can be combined in a menu voice.

The description of the runtime state of a graphical object has to be translated into a description of the runtime state of the corresponding vocal object and vice versa. For example, if users select an option in the graphical interface they can listen its result through the feedback of the choice in voice menu and vice versa, if users press a

particular number of keypad they can see the radio button corresponding the same option selected in the graphical interface.

In graphic-to-graphic migration, the runtime state is sent to the target device in a form adapted to its resources. Applying the state to the page is the task of the migration client running on the device. The same technique can be used in performing vocal to graphic migration, but not for graphic to vocal because a typical vocal device has very limited capabilities. Hence, when the target has no computational capability other than handling phone calls, all the work must be performed on the server side. Once the vocal target page has been identified, a new temporary page is created. Such a page is a copy of the target one, plus the suitably adapted runtime state. In this way, the original page remains available on the server for access by other users and the modified copy is removed when the target platform ends the call.

5 The Migration Service Architecture

The scenarios introduced in Section 3 can be supported by a migration service that allows trans-modal migration to be activated either by user request or automatically when the environment conditions require it. We define the first type as *on demand migration* and the second *automatic migration*.

In *on demand migration*, the user explicitly asks for the application to migrate, specifying the target device. In *automatic migration*, the system must check the environment conditions like mobile device battery energy level and device proximity, in order to decide if migration is needed and to select the target device when more than one fit the predefined requisites.

In [1] we proposed a solution to support on *demand migration* involving devices supporting graphic Web browsers. In this work, we improve the migration service adding a modal migration from graphic to vocal browsing and vice versa. We also add the *automatic migration* service by changing the runtime context manager module and introducing client devices classification and localisation mechanism. In our previous work, the runtime state of the migrating page was collected on the client side and sent to the server only when the user decided to migrate the application. In the runtime state there is the result of the user interactions (selected elements, values entered, …). In the new solution, we keep updating the server-side data structure, describing the runtime state of the application on the client. In this way, the server does not have to query the client for its runtime state, in particular, when migration is triggered because a previously available device becomes unavailable. Otherwise, it would not be possible to retrieve the runtime context of the application running on it. The new solution for the state management is discussed in 5.1.

The other important new feature is the introduction of the client device classification and localisation. Devices are classified in terms of features that guide the selection of the target device for *automatic migration*. In particular, they are considered as part of an environment as described in 5.2 Activation of the application on the target device has also been improved, in order to enable modal migration, which was not previously supported. The new solution is discussed in 5.3.

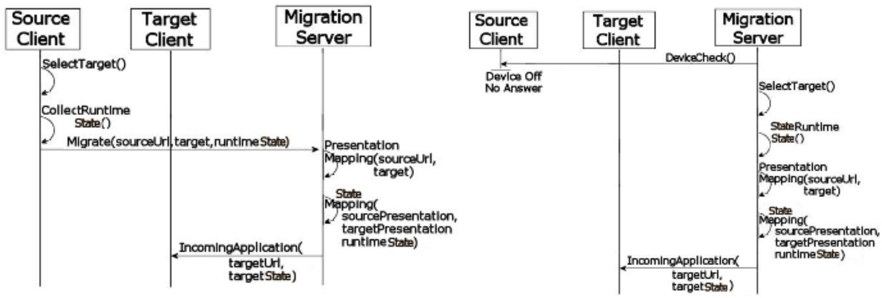


Fig. 1. Left: old migration solution, right: new migration solution

The service relies on a server machine that stores the migratable applications, as well as their logical descriptions and the mechanisms to perform device migration. The server is initialised by building correspondences between tasks and the user interface logical elements, and between the logical description of interface elements and the objects used for their implementation. Such operations are performed for all migratable interfaces and for each one of their device specific versions (see section 4). The time required for the initialisation phase increases with the number of supported applications and their complexity. However, this operation is performed only once at start-up and considerably increases the speed of runtime migration. Users who want to access the service have to load the migration client from the server onto their device. This operation allows the server to identify the devices available for migration and also enables the user device to send migration requests and work as a target for an incoming migrating interface.
Summarising, the main aspects of this improved version of the migration service are: state management, device management and target interface activation.

5.1 User Interface State Management

When a new device enters the migration service, a state collection module is activated on the client side and a corresponding one is created on the server side.
Generally speaking, a server can not access directly information inserted by users on client devices, until they are submitted. In performing migration, we need what has been inserted in the page shown to the user, for this purpose clients can provide useful support in the runtime state collection.
Any time the user interacts with an element of an interface, the client module catches the generated event and immediately sends the new state of the object to the server. The captured events relate to actions, such as objects selection and text insertion. The server keeps a description of the runtime state of the pages loaded on the client, and updates it at each new message received by the client.
When a migration request has to be served, the server analyzes the description of the runtime state, associated to the client from which the interface has to migrate to retrieve the URL of the last page visited by the user and the runtime state of each object of the interface as it was when migration was requested (or triggered).

The last visited page URL is used in the process of target page retrieval and the runtime state of the source page is elaborated to be adapted to the retrieved target page (see section 4).

5.2 Device Management

When asking for *on demand migration*, the user specifies which device has to be the target. In this case, the only information concerning the target device that is necessary, is a description of its type and its supported features. In *automatic migration*, the target is selected by the server among the devices registered to the service according to their features, settings and location.

- *Features*. When a device accesses the migration service, the server recognises its platform type like mobile phone, PDA, desktop, vocal and features such as the screen size, browser supported, etc. In particular, client devices are also recognised as mobile or stationary.
- *Settings.* When the users start the client migration module, they have to specify if the device has to be used as a personal or shared device. A device is shared when more than one user can access it, while it is personal when only the owner can use it. The availability to accept incoming migrating interface has to be declared. Users can also register to the service, specifying more devices that must be considered as potential target for migration. Such devices are those that cannot load a migration client, but can be activated directly by the server, in particular they can be fixed or mobile phones.
- *Location.* The server must keep track of the position of each active client. Devices are considered near, when they are inside the same environment. An environment can be a room, when we consider a building or a car when we consider the user moving outside. The current environment is mainly detected through the use of WLANs and infrared beacons. Stationary devices such as desktop PCs, are statically considered into a specific environment that can not change until the device is turned on, while mobile devices are subject to change position frequently and their position is kept updated.

When selecting a target device for automatically triggered migration, the server considers all the devices being in the same environment in which the source device is that are available to receive incoming applications. In order to select the final target device, among a set of available candidates the migration server analyses the interaction capabilities and energy supply matters of the available devices. For example, we can think of a user interacting with a vocal application through his mobile phone, while reaching his desktop PC and having his PDA turned on in a pocket. The mobile phone is losing battery power and turns off, the application must migrate, and both the PDA and the desktop are close enough to the user. In this case, the desktop is selected as the target device, because a PDA could also be affected by energy supply problems and offers less interaction facilities than the desktop.

Checking the environment, the migration assigns priority to the devices registered as personal and that can be automatically activated. In case the user has for example a fixed or mobile phone in his device list, the server can make the phone ring migrating the application to one of them as soon as the user answers the call.

5.3 Target Interface Activation

There are two different modalities used to activate the interface application on the target device. In case of vocal to graphic migration, the target is required to run the client migration module. Once the server has calculated the URL of the page to be loaded on the target and adapted the corresponding runtime context, all information is coded in a formatted string and sent to the client module running on the target. The client module extracts the URL from the string, loads it into a Web browser window and also extracts the runtime state and applies it to the new page.

In case of graphic to vocal migration, if the host platform corresponds to a fixed or mobile phone then the server is instructed to send a phone call to the appropriate target, indicating which presentation has to be activated on user phone answer and how the vocal interpreter has to run the target presentation applying the runtime context obtained by the migration process.

Migrating from a modality to another one goes far beyond a simple one to one mapping among the pages of the two different versions.

The graphical interfaces do not translate well into speech interfaces for a number of reasons. For instance, graphical interfaces do not always reflect the vocabulary that people use when talking to one another in the application domain. Another important consideration concerns the information organization. In fact, presentations that work well in the graphical interface can fail in speech implementations. Reading exactly what is displayed on the screen is rarely effective. Likewise, users find it awkward to say exactly what is printed on the display. Therefore, it is necessary to analyse the logical description of the application to obtain a graphic to vocal mapping and vice versa, based on the supported task sets.

6 The Multimodal Restaurant Booking Application

In this section we introduce the Multimodal Restaurant Booking Application, a sample application built on the basis of one scenario described in Section 3. In the application the user can choose a restaurant in a specific area of the city. After selecting the Mermaid restaurant, the user fills in the form for booking a table. Let us imagine that he has filled in the first three fields and has selected menu type and then realises that it is getting late, so he decides to continue his booking by phone with the voice system in the car.

The first step is performed by the migration service in order to identify the voice presentation most similar to the source graphical presentation. Then, the migration server accesses the Abstract User Interface of the graphical interface and retrieves the

presentation corresponding to the migrating page. At this point, the set of tasks performed by the presentation, and therefore supported by the migrating page, is identified and used by the mapping algorithm in order to find the right target presentation. In the graphical application the set of tasks are composed of: *provide name*, *provide e-mail*, *provide date of reservation*, *provide time of reservation*, *provide number of people*, *select preference seating*, *select menu type*, *provide special request or comments*, *send reservation*, *and cancel reservation*.

During the mapping the migration server compares the task set of the source presentation (graphic) with the task set of the target (vocal) and identifies the most similar abstract presentation of the vocal abstract interface. During this step it may happen that some tasks supported by the source platform cannot be supported or can be performed through different interaction techniques. For example, the sample application does not support the task "*Provide special request or comments*" in the voice platform, because it would encumber the vocal interaction as it is not an essential task for booking.

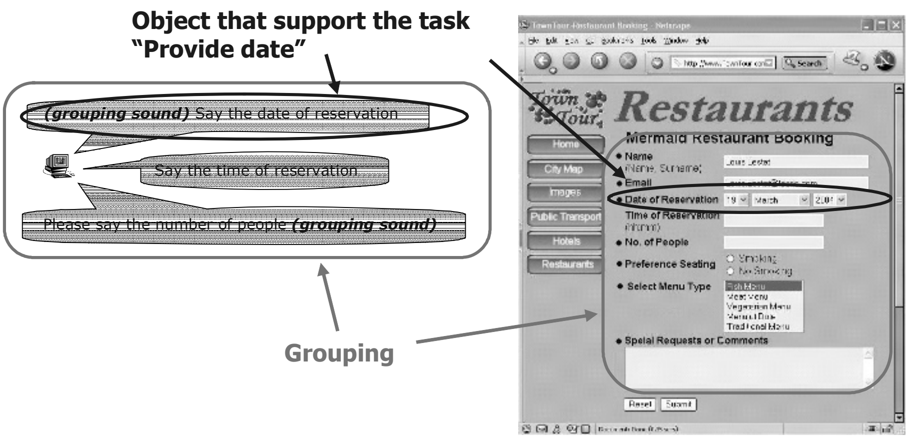


Fig. 2. Presentations of some booking information in different modalities.

Another example is the different method used for supporting the task "*Provide Date of reservation*". In the desktop interface it is implemented by three pull-down menus (day, month and year) while in the vocal system it is accomplished by a vocal input request to the user for the date of reservation without indicating any potential choice (see Figure 2).

In the example, the migration server identifies three vocal abstract presentations containing the same number of tasks of the source presentation. One task is not supported in the vocal application (*Provide special request or comments)*. The first presentation requests the user's name and e-mail, the second presentation requests the reservation date, time and the number of people, and the third presentation requests seating preferences, the type of menu and confirms or deletes the reservation.

It is also interesting to notice the different techniques adopted to combine interactors. For example, in the graphical interface the grouping operator is obtained through an

unordered list, whereas the vocal interface uses a sound to delimit the grouped elements.
The second step allows the migration server to select the presentation that contains the object implementing the last task performed by the user on the target platform.
During this phase, it is important to consider that the vocal channel serialises interactions, while they can be performed concurrently on a visual channel. Accordingly, once the presentation has been identified, the migration server checks if all the previous tasks have really been executed. If a negative response results for any tasks, they are performed first and then the dialog carries on from the task last executed in the source device.
In the presented example, composed of three vocal presentations, the last executed task is *select menu type* and is included in the third presentation; the first presentation, which asks for the user name and the e-mail, has been performed, while the second and third presentations were not completed. In this situation, the dialog starts with the first task of the second presentation (*provide time reservation*) and skips the tasks that have already been executed through the graphical interface (see Figure 3).
With this solution, the data previously inserted in the form by the user are not lost, and can be listened to in a feedback message of the last presentation.

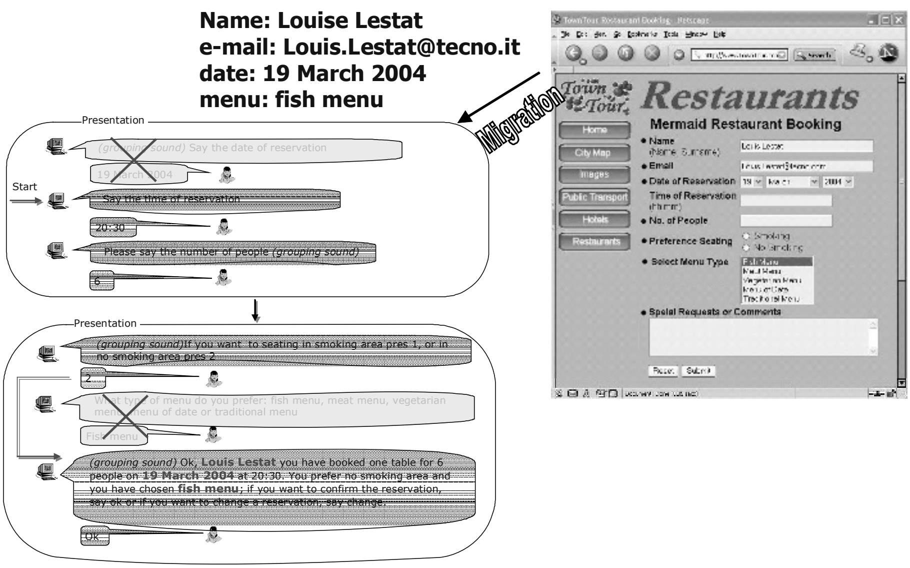


Fig. 3. Example of Migration scheme

7 Conclusions and Future Work

We have presented a new solution to obtaining migrating interfaces that can be either initiated by the user or automatically triggered by the system when environment conditions require. Moreover, we have also added the possibility of interaction

modality changes during migration. In particular, we have addressed graphic to vocal migration and vice versa.

At this stage, our prototype of migration service supports migration of interfaces implemented by XHTML, XHTML Mobile Profile and VoiceXML developed with TERESA. We will soon support also multimodal interfaces implemented in languages such as X+V.

Further studies will address the improvement of the migration service in order to support Web interfaces developed using other tools as well. This further issue will require a different kind of interface analysis: we plan to use tools for reconstructing a logical description of the pages at runtime. The extension to such interfaces is a main goal for our future work.

Another topic for future work is the support of multimodal distributed migration, in which a user interface migrates in such a way to carry on interaction through multiple devices.

Acknowledgments. We gratefully acknowledge support from the IST EU R&D CAMELEON project (http://giove.cnuce.cnr.it/ cameleon.html) and would like to thank our colleagues in the project for useful discussions.

References

1. Bandelloni R., Paternò F., Flexible Interface Migration, Proceedings ACM IUI 2004, pp.148-157, ACM Press, Funchal, 2004.
2. Coutaz J., Balme L., Barralon N., Calvary G., Demeure A., Lachenal C., Rey G., Bandelloni R., Paternò F., the CAMELEON Run Time Infrastructure for User Interface Adaptation, October 2003, CAMELEON Deliverable D2.2.
3. de Sousa, J., Garlan, D. Aura : an Architectural Framework for User Mobility in Ubiquitous Computing Environments. IEEE-IFIP Conf. 140 on Software Architecture, Montreal, 2002.
4. Ferreras, C, G., Mancebo, E, D., Payo, V, C.. From HTML to VoiceXML: A First Approach. In Proceedings of TSD 2002, LNAI 2448, pp 441-444, 2002.
5. Mori, G., Paternò, F., Santoro, C. “Tool Support for Designing Nomadic Applications”, Proceedings ACM IUI’03, Miami, pp.141-148, ACM Press.
6. Nichols, J. Myers B. A., Higgins M., Hughes J., Harris T. K., Rosenfeld R., Pignol M.. “Generating remote control interfaces for complex appliances”. Proceedings ACM UIST’02, pp.161-170.
7. Pérez-Quiñoes, M, A., Capra, R, G., Shao, Z.. The Ears Have It: A Task by Information Structure Taxonomy for Voice Access to Web Pages. In Proceedings of INTERACT 2003.

Web Page Transformation When Switching Devices

Bonnie MacKay, Carolyn Watters, and Jack Duffy

Faculty of Computer Science, Dalhousie University
Halifax, Nova Scotia B3H 1W5 Canada
{bmackay, waters}@cs.dal.ca, jack.duffy@dal.ca

Abstract. With network and small screen device improvements, such as wireless abilities, increased memory and CPU speeds, users are no longer limited by location when accessing on-line information. We are interested in studying the effect of users switching from a large screen device, such as a desktop or laptop to use the same web page on a small device, in this case a PDA (Personal Digital Assistant). We discuss three common transformation approaches for display of web pages on the small screen: Direct Migration, Linear and Overview. We introduce a new Overview method, called the Gateway, for use on the small screen that exploits a user's familiarity of a web page. The users in an initial study prefer using the Gateway and Direct Migration approach for web pages previously used on the large screen, despite the common Linear approach used by many web sites.

1 Introduction

With network and small screen device improvements, such as wireless abilities, increased memory and CPU speeds, users are no longer limited by location when accessing on-line information. Rather, small screen devices have enabled users to access information, in particular the Internet, from any location with relative ease. In 2002, ComScore Networks Inc. [5] reported that 9.9 million American adults use their PDA (Personal Digital Assistant) or cell phone to access the Internet with news sites being the most commonly accessed web pages. With multiple devices, users can move between these devices while accessing the same information. Users could use a web page on their desktop at the office and use the same information on their PDA while commuting home.

Despite the technical and bandwidth enhancements, PDAs are restricted by the small size of the screen that limits the amount of information that can be displayed at one time. While some research on the effects of different line lengths for reading has found that the limited screen size has little effect on comprehending information, it has been shown to influence reading rates [8],[9]. The small screen can also affect the display of many common web information structures, such as graphs, tables and forms. Using the small screen to effectively access information is further influenced by the very nature of PDA's: their portability. Users using PDAs "on the go" subject themselves to noisy environments with the high probability of interruptions and movement [11]. Similarly, this portability could negatively affect accurate selections on the screen and entering information.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 228–239, 2004.

There are two broad approaches for displaying web pages within the small screen constraints of PDAs. The first approach is based on generating static web pages specifically designed for small screen devices. The second approach utilizes some form of automated transformation of the original large web page. The obvious advantage of an automated transformation is the increased pool of accessible web pages for PDA users. However, many current automated transformation options do not consider features such as user task, familiarly with information, web page layout and mobility of the user, and their impact on the usability of the resultant transformed page.

In this paper, we will discuss three approaches to transform web pages to the small screen. We introduce a new method of automatically transforming existing web pages, called the Gateway, for use on the small screen that exploits a user's familiarity with the page to reduce transformation volatility. Transformation volatility results from changes to the look, design, layout and even content when using the same web page on different devices. Finally, we will describe a user study comparing three different display approaches in this context.

2 Web Display Approaches for Small Screens

Many web sites provide a small screen version of their pages for their PDA users. Images may be removed for a text-only version or reduced in size to fit the screen. Font styles and sizes may be changed and reduced. Often the layout, display and sometimes even the content of the original web page are transformed to fit within the constraints of the small screen size. Internet browsers are now available that are better suited to web browsing on the small screens. For example, Windows CE IE has added word wrap and allows users to change the font size to better fit web pages within the screen constraints. Web page transformation, whether at the site or at the browser level, can be divided into three broad transformation categories: Direct Migration, Linear and Overview [14].

2.1 Direct Migration

For Direct Migration, there are no transformations made to the original web page. While this approach does not require human or system intervention, it does require more effort to navigate the page by the users. Users must navigate using both vertical and horizontal scrolling which can cause user frustration and reduce the usefulness of the information on the page as only a small part of the page is visible at one time [1], [9], [12]. Despite the negative points associated with this approach, Direct Migration does provide ready access to most web pages. It can be considered the default transformation for pages without small screen versions. Although browser upgrades have improved web page access on the small devices, these browsers are still limited by the inherent design structure of web pages, such as tables used for formatting and frames.

2.2 Linear Transformation

This approach is used by many web sites, such as news sites, for their users of small devices. Sites create their own Linear versions or use a service such as Avantgo (www.avantgo.com) or Usable Net (www.usablenet.com) to transform the main site into this format. The layout of information from the main web site is changed to a long linear list that fits within the width constraints of the small device. Images may be reduced or even omitted thereby decreasing bandwidth and download time. Content may be changed or reduced using techniques such as summarization [3] or even removed. The main benefit of this approach is that horizontal scrolling is no longer needed, although vertical scrolling may increase substantially. Users navigate by vertically scrolling and clicking to expand links, such as headlines to retrieve more detail.

2.3 Overview Transformation

This form of transformation provides users with an overview of the original web page. Overviews, such as focus + context [15] and Fisheye views [10] have been used successfully on large screens to display large and complex data sets. This approach has been adapted for use on small devices to display large web pages within the constraints of the small screen [2], [4], [17]. For example, the West Browser uses flip zooming [2] that adapts the fisheye approach for the small screen by dividing a large web page into a hierarchy of individual pages or cards that users can flip through. Each card contains up to seven objects that are a representation of information, such as a thumbnail image or text that users can expand for more detail.

The advantage of an Overview approach is that part or the entire layout and, for the most part, content remains the same as the original web page. As well, scrolling may be reduced or even eliminated. The disadvantage of this approach is that by shrinking the original page, readability becomes an issue requiring creative solutions. For instance, the Thunderhawk browser (www.bitstream.com) uses a landscape view to increase the screen width and a special font that replaces the original web page's font at a considerably smaller size while maintaining the readability of the font. While this helps maintain the layout and consistency of the original web page, users often still need to scroll both vertically and horizontally to view the page on the small screen.

3 Design Motivations and Issues

We first explore the usability issues associated with using web pages on the small screen, including web page layout, familiarity with the web page, user task and mobility, and their impact on both the usability and suitability of web page transformation on small devices. We then introduce a new method of automatically transforming existing web pages, called the Gateway for users who are already familiar with web sites.

3.1 Usability Issues

The first usability issue is *web page layout*. The success of automated transformed pages largely depends on the quality of the original web page. Watters et al [16] generalized the layout of web pages into two broad categories: Broadsheet and Linear. Broadsheet web pages tend to be organized into columns with a combination of images and text, similar to a glossy brochure. Many news sites use this approach. Linear web pages tend to contain more text and require scrolling to read. These pages may be very simple with little navigation features or may contain navigation options using a side or top menu bar. Web pages authors have access to simple usability guidelines to improve the overall quality of web pages. However, many pages still vary on many characteristics such as page length or scrolling, color combinations, and font sizes.

A user's *familiarity* with a web page is the second usability issue. When a user first uses a web page, they establish a mental model of the page based on the structural organization of the information, such as visual cues, layout and semantics [1],[7],[15]. A primary objective when transforming a web page for different devices is to minimize the user effort in re-establishing the existing mental model of the original page. Danielson [7] introduced the concept of transitional volatility and described two ways the web is volatile: web sites can change over time and within sites users can experience different navigation structures. Danielson [7] found that a highly volatile session increased disorientation and decreased user navigation abilities. When users switch between devices to use the same web page, this introduces a new type of volatility: transformation volatility [16]. Transformation volatility is a measure of change to navigation, layout, content and readability from one device to another. When a user accesses a web page on a desktop and uses the same web page on their laptop, the transformation volatility is small. But when the user uses the same web page on their PDA the transformation volatility is substantial. Our goal is to minimize the transformation volatility for users switching between different screen sizes to access the same web pages.

The *type of tasks* that the user engages in is another usability issue. Users access the web for different reasons at different times. We have identified five web-based tasks that users frequently engage in: re-finding information, finding new information, comparing information, reading information and general browsing. That is, users may need to re-find information that they have already seen. As well, users may need to find specific information that they have not seen before, e.g. a student looking for references for a paper. Users may want to compare information or details, such as airline prices or dates which could involve looking up information on one page or it could involve going between pages. Users may want to read the web page, such as a news story or journal paper. Finally users may just be browsing the Internet. This browsing may be for general interest, for example planning your next vacation, or it may just be the act of randomly choosing web pages and following links with no particular goal.

The last usability issue that we have identified is *mobility*. Different factors impact the user experience when users are mobile using their PDA to access the web. Some of these factors are external to the experience, such as noise, distractions and movement. While we can not influence these factors, they have an impact on the user. When users are moving, either physically themselves or while on the move, such as being on a bus, scrolling and clicking using the stylus may become difficult.

Distractions, noise and movement can all affect the user's ability to read and concentrate [11], especially if the user is also trying to navigate the web page in the small window using the stylus. A small screen version that reduces the necessary scrolling and clicking would be beneficial in such times, such as an Overview transformation.

3.2 The Gateway

The Gateway is a new Overview transformation prototype designed specifically to minimize the transformation volatility for users who switch between devices to view a familiar web page. The Gateway differs from previous focus + context models in that it provides an exact reduced replica of the large screen web page while maintaining a consistent distortion. Users navigate the Gateway by selecting individual sections, either by clicking or by rollovers, on the web page that are expanded and superimposed over the overview (Fig. 1). Users can then make selections on the section, such as choose a menu item or follow a link, as they would on the large web page. The Gateway is similar to Microsoft's adaptive viewing approach [4], but the Gateway provides a zooming capacity more consistent with the focus + context approach.

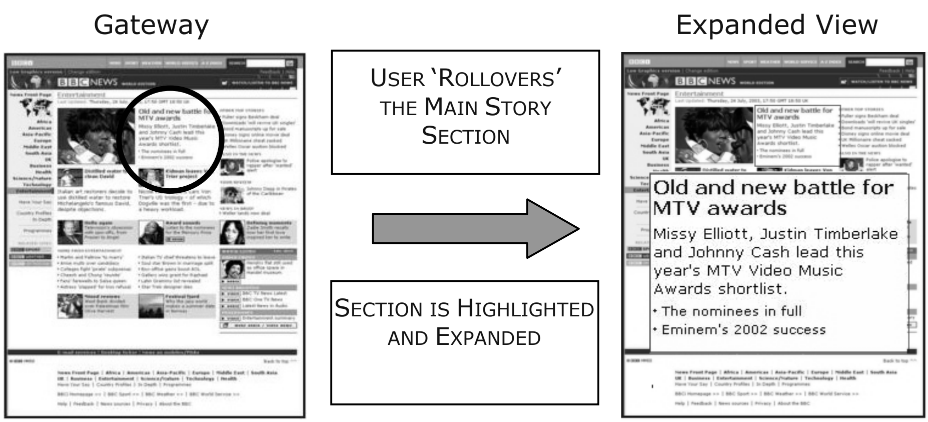


Fig. 1. The Gateway

The Gateway provides a thumbnail style representation of a large web page at a pixel size of 240 by 320. Research on the use of thumbnails for web tasks has found that thumbnail representations of web pages are useful visual memory aids for users that improves user recall [13],[18]. Kaasten et al. [13] found users had a high recognition of web pages when thumbnails had a pixel size as small as 208 by 208. As well, the Gateway maintains the spatial location of the original page that has been shown to help users develop a mental model to make sense of the organization of a page thereby helping users to remember the location of features on the page [6].

3.3 Automatic Transformation of Web Pages

Clearly, there may not be one best automatic transformation for all web pages. The best transformation for mobile devices may depend on the original large web page layout, familiarity of a web page, a feature set of user tasks, and level of user mobility. We conducted a study to examine these features for users familiar with a web page. Users rated each small screen version for five tasks (finding, re-finding, reading, comparison and browsing) based on their experience using each version. The web page chosen for the testing was a news site with a Broadsheet layout. A Linear web page layout was not used in this study but will be included in further testing. Finally, users used two of the small screen versions while moving around to gage the mobility of each approach.

4 User Study

4.1 Methodology

We had ten computer science graduate students participate in the study, ranging in age from 25 to 55. There were five female and five male participants. It was a within subject study, where each participant viewed three small screen versions in a different order on the large screen and used the Gateway and Linear version on the PDA in alternating orders. The shortest time to complete the study was about 40 minutes, while the longest session was about 50 minutes. Testing was conducted using both a desktop computer with a 15” monitor and a Toshiba e750 Pocket PC using the BBC news site Entertainment section that was downloaded to a local machine. Users were tested on small screen transformations based on the regular sized web page using three different interfaces (Fig. 2): Direct Migration, Linear and the Gateway.

Fig. 2. Three Small Screen Transformations

The Direct Migration was the actual large web site shown on the small sized screen that required users to scroll both horizontally and vertically to navigate the page. The Linear transformation was BBC’s own actual linear textual version that had changes in the layout from the large web page and some stories were omitted. The Gateway

used a reduced replica of the actual BBC web page with rollovers to identify the different regions on the page.

4.1.1 Hypotheses

i) Users who have previously used and are familiar with a web page will prefer using the Gateway to the Linear and Direct Migration transformation approaches.
ii) The Gateway will be preferred for tasks of re-finding information and comparing details on a web page.
iii) Linear transformation will be preferred for finding and browsing information on never seen web pages.
iv) Direct Migration will be the least preferred version for all tasks and general user satisfaction.

4.1.2 The Study. The study consisted of two parts. The first part, a Comparative Task Completion asked users to perform a task using and comparing three small screen versions that were all displayed at the same time on a desktop. First, users became familiar with the large BBC web page on the desktop by opening stories, reading the headlines and using the menu items. Before the actual testing, users performed the same manipulation check (e.g. go to the sports section, find a main story, etc.) to ensure each user had a minimum level of familiarity with the page. Users were asked to choose three stories on the large page that they would recommend to friends. At least one story could not contain a picture or an image. The large version of the web page was then closed and replaced with three small screen versions (Direct Migration, Linear and Gateway) in random order (Fig. 2). Users did the same manipulation check on each of the small screen displays after finding one story on each version. Once finished, the users were interviewed and asked a set of questions relating to their task experience and preference.

The second part, Mobility Feedback, had users actually move around while using the PDA. Users were asked to use both the Linear and Gateway version on the PDA to locate the three stories they found on the large BBC web page. We did not include the Direct Migration approach as we were concentrating on and comparing the Gateway's Overview approach with the commonly used Linear approach. Once users found the stories using both versions, users were asked about their experience of using each interface on the PDA and some general design questions regarding the Gateway.

5 Evaluation and Results

5.1 Part I: Comparative Task Completion

5.1.1 User Preference Results. Users were asked to rank all three versions based on four user preference questions: the fastest to find the story; the easiest to find the story; the most intuitive to use and liked using best to find the story. Users ranked each version by giving the 'best' a score of 1, the 'next best' a score of 2, and the

'least preferred' a score of 3. In Table 1, we added the scores of all users to measurethe preference for each question and for an overall user preference score. The best a small screen version could score on an individual question was 10 (10 users times a ranking of 1). The worst score a version could receive on an individual question was 30 (10 users times a ranking of 3). The best overall score a version could receive was 40 (best score of 10 times 4 questions) and the worst overall score was 120 (worst score of 30 times 4 questions).

Table 1. Overall User Preference Scores.

Category	**Gateway**	**Linear**	**Direct**
Thought were fastest on	15	29	16
Easiest to find story	17	28	15
Most intuitive	16	29	15
Liked using	13	30	17
Total	**61**	**116**	**63**

Overall, we found the results quite surprising. We hypothesized that users who were familiar with a web page on a large screen would prefer the Gateway and Linear, and that users would least prefer the Direct Migration approach due the increased effort to navigate the large web page on the small screen. The chi-square test on Table 1 shows that the Gateway is significantly better than the Linear version although not different than the Direct Migration approach. The chi-square equaled 24.2 with 2 degrees of freedom which is significant as a one tailed test at $p < .005$. As well, Direct Migration is significantly better than the Linear version. Linear, a common transformation used by many web sites had the worst score of 116, only four points from being ranked the "worst" version. In fact, users often referred to the Linear version as "hateful", "annoying" and the "worst". One point that influenced the Gateway ranking which was noted by many of the users was that they did not understand at first how to select from the expanded section from the rollovers. When the rollover occurred, users had to click once to then select from the expanded selection. A training session could have alleviated this uncertainty.

We had expected that users would find the lack of readability an issue with the Gateway and had already considered a design for a revised Gateway to improve readability. Particularly interesting was that the users in this study understood readability differently than strictly being able to read the font size. Users found the Gateway readable because it maintained the same layout as the large page and they could expand the sections with the rollovers. Only one user noted that the lack of readability for the Gateway negatively influenced the ranking preference for it. Despite both the readable font size of the Linear version and that it only requires vertical scrolling, some users chose the Direct Migration as their first preference in some categories. The main issues with the Linear version was that content and layout had changed and while finding the main pictures or top stories was not difficult for some, all seemed to have difficulty finding less obvious stories and menu items.

5.1.2 Task Results and User Comments. Users were asked to rank each version based on five different tasks commonly performed on news web pages: reading astory, find a never before seen story, re-find an already seen story, compare details

between stories and general browsing. Once again, users ranked each version by giving the ‘best’ a score of 1, the ‘next best’ a score of 2, and the ‘least preferred’ a score of 3. In Table 2, we added up the scores of all users to measure the preference for each question and for an overall best task-based score. The best a small screen version could score on an individual question was 10 (10 users times a ranking of 1) and the worst score was 30. The best overall score a version could receive was 50 and the worst overall score was 150.

Overall, users ranked the Gateway the highest with a score of 73, which is much more decisive gap over the other two versions. We had hypothesized that users would prefer using the Gateway over the Direct Migration and Linear approaches for familiar web pages. Using chi-square we found that the Gateway is significantly better than both the Linear and Direct Migration approach for performing tasks, where the chi-square is 10.24 with 2 degrees of difference which is significant as a one tailed test at $p < .01$.

Table 2. Task Scores.

Task	**Gateway**	**Linear**	**Direct**
Reading a story	11	12	30
Find a new story	18	19	23
Re-find already seen story	14	29	17
Compare details	14	21	25
General browsing	16	22	22
Total	**73**	**103**	**117**

Unlike the readability of the full web page, *reading a story* refers to the actual reading of a news story expanded from the original web page. The Gateway used the exact same story layout as the Linear version with the menu items deleted from the top and bottom of the page. Since the Linear and Gateway versions had the same story and layout of the story, this was the only category (question) we allowed users to give a tie between versions. Direct Migration had the worst score. Users had to scroll both vertically and horizontally to read the story. Users noted that it was “horrible” and “annoying”.

It was interesting that the Gateway ranked so well to *find new stories*. We had first thought the Linear version would be preferred for a task to locate a new, never seen story because users could read the content by navigating in one direction. Still, four users ranked the Linear first and three users ranked the Gateway first. The main problem that some users noted with the Gateway was that they would be unsure if they had missed something on the page, in that they would not know if they had “expanded all the boxes”. Users noted that the Linear version would be fine for main stories or information located at the top of the page but that it would be slower to locate other stories as it lacks important visual cues (such as colours and page layout).

Overall, users felt that the Gateway would be best to *re-find stories* already viewed on the large screen, followed closely by the Direct Migration version. Seven users ranked the Gateway first, with only one user ranking it last. Linear was ranked by all but one user last. Interestingly, we had thought that users would find the Direct Migration version, although exactly the same as the large screen, not very useful as one could only see a small portion of the page at anyone time. Still, these results were

as we expected. When users have already viewed and located information on the large screen, they can transfer their existing mental model of that page to the smaller version and re-find the same information easier using the same layout than a different layout.

To *compare details* users were given the same demonstration using the large web page for a task that required going between two different but related stories from the main page. Users then tried the same task using each of the three small screen versions before ranking each version. Overall, the Gateway was ranked the highest with a score of 14 with a considerable gap between the next highest score for Linear with 21 points, followed by the Direct Migration with 25 points. Seven users ranked Gateway as first for this task, with one ranking it last. Six of the users ranked Direct Migration last for this task. One user noted that the Gateway allowed them to go between the stories very easily. Many users noted that being familiar with the page before conducting the comparison made a difference in their rankings. One user stated that the Gateway "gives a birdview. It is very easy to go where you want especially if you know where to go". Users noted that getting to the stories for a familiar site with Direct Migration "wasn't so bad", but then reading the actual stories to get the details to compare was difficult.

General browsing includes viewing never seen pages, which makes the results on this category very surprising. We had believed that the Gateway would do well for tasks using a familiar web page but believed that Linear would do better for an unfamiliar web page, once again due to the readability factor. Still, users quite decisively ranked the Gateway the highest with a score of 16, although only five users gave the Gateway a ranking of 'best'. One user noted that the "Gateway is good because you can see the relevance of importance of information with the overall structure".

5.2 Part II: Mobility Feedback

Users commented that they liked using the overview of the Gateway to find the stories and found the Gateway to be more navigation based. It should be noted that the Gateway version on the PDA was slightly different then the version on the desktop due to shortcomings of the actual browser on the PDA; it did not use rollovers but required users to click to expand the specific sections on the overview. Only one user said that they preferred not having the rollovers on the mobile Gateway version, while others noted that they preferred the rollovers. With the Linear version, users stated that they did not like the reformatting of the layout from the original page and found it easy to get lost. A user noted it "was frustrating because I knew where to look if I had been using the other version [the Gateway]." Users also commented on features that they felt as important for web use on the small screen that included completeness and full access to information, consistent layout, readability and no horizontal scrolling.

5.3 Design Feedback

We asked users to provide feedback on a revised version of the Gateway to help with the readability issue associated with the Gateway. We adapted the existing Gateway

prototype to enlarge titles and expand keywords from the story headings. So while descriptions under headlines or pictures are still unreadable, users can read the actual sections on each page and in addition to having the visual cues from the original web page, can have word cues. Woodruff et al [18] found that thumbnails enhanced with text performed as well or better then plain thumbnails. Overall, users found this to be a positive improvement. Readable headings could allow users to be more selective with the rollovers and help users quickly scan the page.

6 Conclusion and Future Work

Despite the prevalence of Linear versions for small screens, we have shown that users prefer small versions of familiar web sites that are more closely related to the mental model of the larger version and that users preferred the Gateway for web related tasks. Users generally found that the change in navigation structures, layout and content from the large web page to the small Linear version caused confusion and disorientation, especially for re-finding information and comparing information. This was evident when users were both stationary using the desktop version and mobile using the PDA. The Linear versions may be advantageous when users are restricted by bandwidth and processing power or have very small screens, such as mobile phones.

We are ready to perform user studies to compare the efficiency and effectiveness of the Gateway transformation model with the Linear approach using both Broadsheet and Linear web page layouts. We will test users on simple lookup tasks of re-finding and finding new information and on a more complex comparison task. We will compare the results of users who first view a web page on the large screen then switch to the small screen using both the Gateway and Linear model. We will also test users using previously unseen web pages on the small screen device using both the Gateway and Linear model. We speculate that the Gateway will perform better for web sites previously viewed on large screen devices; however, similar to this user study we may have underestimated the impact of the graphical layout on the large screen.

References

1. Albers, M. & Kim, L. User Web Characteristics Using Palm Handhelds for Information Retrieval. In *Proc. of IPCC/SIGDOC Technology and Teamwork*. Cambridge, MA (2000), 125-135
2. Björk, S., Holmquist, L.E., Redström, J., Bretan, I., Danielsson, R., Karlgren, J. & Franzén, K. West: A Web Browser for Small Terminals. In *Proc. of ACM Symposium on User Interface Software and Technology*. Asheville, NC (1999), 187-196
3. Buyukkokten, O., Garcia-Molina, H. & Paepcke, A. Seeing the Whole in Parts: Text Summarization for Web Browsing on Handheld Devices. In *Proc. of 10th Internat. WWW Conf*. Hong Kong (2001), 652-662
4. Chen, Y., Ma, W. & Zhang, H. Adapting Content to Mobile Devices: Detecting Web Page Structure for Adaptive Viewing on Small Form Factor Devices, In *Proc. of 12th Internat. WWW Conf.* Budapest (2003), 225-233

5. ComScore Networks. Ten Million Internet Users go Online Via a Cell Phone or PDA (August 2002), from http:// www.comscore.com/press/release.asp?id=87
6. Czerwinski, M., Dumais, S., Robertson, G., Dziadosz, S., Tiernan, S. & van Dantzich, M. Visualizing Implicit Queries for Information Management and Retrieval. In *Proc. of ACM SIGCHI Conf*. Pittsburg, PA (1999), 560-567
7. Danielson, D. Transitional Volatility in Web Navigation, *IT&Society*. (2003), 1(3): 131-158
8. Dillon, A., Richardson, J. & McKnight, C. The Effect of Display Size and Text Splitting on Reading lengthy Text from the Screen, *Behavior and Information Technology*, (1999), 9(3): 215-227
9. Dyson, M. & Haselgrove, M. The Influence of Reading Speed and Line Length on the Effectiveness of Reading from Screen. *International Journal Human-Computer Studies*, (2001), 54 (4): 585-612
10. Furnas, George W. Generalized Fisheye Views. In *Proc. of Conf. on Human Factors in Computing Systems*. Boston, MA (1986), 16-23
11. Jameson, A., Schäfer, R., Weis, T., Berthold A. & Weyrath, T. Making Systems Sensitive to the User's Time and Working Memory Constraints. In *Proc. of 4th Internat. Conf. on Intelligent User Interfaces*. Los Angeles, CA, (1998), 79-86
12. Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K. & Buchanan, G. Improving Web Interaction on Small Displays. *In Proc. of 8th Internat. WWW Conf.* Toronto Canada, (1999)
13. Kaasten, S., Greenberg, S. & Edwards, C. How People Recognize Previously Seen WWW Pages from Titles, URLs and Thumbnails. In X. Faulkner, J. Finlay, F. Detienne (Eds) *People and Computers XVI (Proc. of Human Computer Interaction)* (2002), BCS Conference Series, 247-265, Springer Verlag
14. MacKay, B. & Watters, C. The Impact of Migration of Data to Small Screens on Navigation, *IT&Society*. (2003), 1(3): 90-10
15. Spence, R. *Information Visualization*. (2001) ACM Press: New York
16. Watters, C. & MacKay, B. Transformation Volatility and the Gateway Model for Web Page Migration to Small Screen Devices. In *Proc. of Hawaii Internat. Conf. on System Sciences*. Big Island, Hawaii (2004)
17. Wobbrock, J.O., Forlizzi, J., Hudson, S.E. & Myers, B.A. WebThumb: Interaction Techniques for Small-Screen Browsers. In *Proc. of ACM Symposium on User Interface Software and Technology*. Paris, France, (2002), 205-208
18. Woodruff, A., Faulring, A., Rosenholtz, R., Morrison, J. & Pirolli, P. Using Thumbnails to Search the Web. In *Proc. of ACM SIGCHI Conf.*, Seattle, WA (2001), 198-205

Mobile Context Aware Systems: The Intelligence to Support Tasks and Effectively Utilise Resources

Russell Beale[1] and Peter Lonsdale[2]

[1] School of Computer Science
r.beale@cs.bham.ac.uk
[2] School of Engineering
p.lonsdale@bham.ac.uk
University of Birmingham
Birmingham B15 2TT UK

Abstract. The complex usage of mobile devices coupled with their limited resources in terms of display and processing suggests that being able to understand the context of the user would be beneficial. In this paper we present a model that describes context as a dynamic process with historic dependencies. We also describe software architecture to support this model, and evaluate its effectiveness in a mobile learning scenario. Preliminary results from our evaluation suggest important issues for consideration in the continuing development of context aware systems and interfaces, including the need for appropriate representation of contextual data to the user, and maintaining a balance between effective support and intrusion.

1 Context Awareness and Mobile Computing

PDAs, mobile phones, and laptop PCs are used by a variety of users in a variety of different environments. Getting hold of information about the user and their environment and then putting it to good use lets us provide timely support for user activities and allow the user to maintain their attention on the world around them. Context is important because it allows us to make use of the environment in a way that supports the user. Current Nokia mobile phones have a set of modes that determine ring volume, text message notification and suchlike, so that different profiles can be chosen for outdoor use, or when in a meeting, allowing the phone to respond most appropriately. Such choices are made manually by the user, but the principle that the different contexts require different actions is the same. For more advanced systems, we can envisage the scenario of a mobile phone that is aware of its user's location, for example, and will not disturb an important meeting. But the same phone, being aware of its user's call list and calendar will permit a call from a pregnant wife. In this way the user themselves forms part of the environment they occupy, and we can use information about the user themselves to further enhance our contextual model.

Having this kind of automated filtering going on is useful for any user. But for a user of a mobile device, this kind of support becomes even more salient. When users are mobile, they are typically involved in other activities not focused on the device

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 240–251, 2004.

itself. For example, visitors to an art gallery would like to maintain their attention on the great works they are admiring, rather than having to perform content searches on their PDA. A tool that can keep up with what is going on with the user and their environment can allow the user to maintain their attention on the world, and can provide timely support for the user's activities. In addition, mobile devices are usually very limited when compared to desktop devices. These limitations mean that steps taken to reduce the quantity and complexity of the information they have to potentially provide allows them to be more efficient and effective. A large amount of interesting work is being done on display approaches to better represent too much information in too little space with too few resources, but effective assistance to that work can be given by systems that reduce the amount of necessary information to be displayed has to be beneficial.

The reason for modelling context is to better understand the user's activity - in our case, of mobile learning. This in turn leads to the design of systems that deliver more appropriate learning content and services. This is useful in three respects. Firstly, it relates the services to time and location, and to the learner's needs and interests, ensuring that they are useful, learnable and enjoyable. Engagement is increased since if the system can provide appropriate information at the opportune time, it can produce a more compelling learning experience. Matching the correct level of information using the most appropriate learning style for a particular user can produce a more effective and enjoyable learning experience. Secondly, it provides for more effective use of resources, which is especially important in the mobile situation, with many different limitations – device processing power, display ability, media capabilities, network bandwidth, connectivity options, intermittent connections – and other aspects of the situation competing for attention. Thirdly, by providing more appropriate information delivered most effectively, it allows the user to focus much less on the technology and more on the actual situation they are in. By producing a system that is responsive to the user's changing attentions and their associated changes in need, mobile learning systems support a much more exploratory, opportunistic and ad hoc approach to learning that potentially suit their users much more.

2 Modelling Context

What is becoming clear is that there are difficulties in implementing context-awareness. Firstly, how do we get hold of contextual information; and secondly, what do we do with it once we have it?

In order to address these issues, we believe that there is a need for a model of context, to facilitate dialogue about what does and does not constitute context for the purposes of enabling context-aware computing, and to enable flexible re-use of context awareness architectures in a variety of scenarios. The problem with this is that 'context' in itself is all encompassing and recursive – it is difficult in light of this to offer a prescriptive model. It is possible to look at the kinds of things that can be used as contextual data, and to build a model from these examples that can help us explore future implementations of context awareness.

2.1 The Technological Approach Versus the User-Centred Approach

A review of the current literature on context awareness research indicates that there is a polarisation of approaches (for recent reviews, see [3, 4]). Much research can be seen to be driven from a technological perspective, focusing on what current devices, sensors, and software platforms can provide in the way of context aware computing. This approach is understandable given the need to consider the technical aspects of how to acquire and use contextual data. However, the focus on this approach is at the expense of another significant perspective: that of the user. In MOBIlearn [1] we are aiming to work from a user-centred standpoint, identifying the kinds of context awareness that might be required by users in specific scenarios of use, and then implementing a context awareness system around them. Our aim is to provide context aware learning experiences in at least three different scenarios, and our experiences so far have taught us that we need a generalised architecture and model for context awareness to enable useful dialogue between project partners. We therefore suggest that there is a need for an increased attention to the user-centred approach and the need for reusable models. In MOBIlearn, we are aiming for a hybrid approach, working from the user-centred perspective, building re-usable models of context, but at the same time maintaining an awareness of technical constraints.

We consider context not as a static phenomenon but as a dynamic process, where context is constructed through the learner's interactions with the learning materials and the surrounding world over time. For mobile learning, there is an essential interaction between the environment, the user, their tasks, and other users. All of these domains provide information in themselves, and can interact with the others in a variety of ways, building a rich model of the current world and hence allowing the system to be more specific in what it offers the user. The environment contains much ambient information, as do the other users in that space. The learning tasks and the user themselves provide a clearer view of what is important to them, whilst all define the knowledge that is useful and available. A simple example clarifies these concepts: environmental information such as geographical position allows us to provide location-specific information, e.g. for a museum. Other user information such as the identification and presence of another person allows us to create a peer-to-peer network for informal chat. But the combination of the two may allow us to determine that the other user is a curator, and we can provide the mechanisms for one to give a guided tour to the other. The combination of models is potentially richer than each on their own.

3 Our Implementation: Context Awareness for Mobile Learning

M-learning, the mobile equivalent of e-learning, is an emerging field of research being embraced by manufacturers, content providers, and academics alike. More and more people are carrying mobile computing devices everywhere they go in the form of PDAs, smart phones, and portable computers. There is something compelling about the possibility of being able to take advantage of these devices to offer new ways of interacting with information. Learners on the move can use mobile devices to take their learning materials into a rich variety of environments – the challenge is how to

make the best use of this environmental richness provide both intelligent content delivery and engaging learning experiences.

The MOBIlearn project aims to produce an integrated architecture for learners with mobile devices. The system includes support for collaborative learning, an adaptive human interface, and context-aware presentation of content, options, and services. We have been exploring the use of context-awareness as part of a larger m-learning architecture to provide an engaging and supportive learning experience in different environments.

The MOBIlearn context awareness subsystem [6], currently being developed at the University of Birmingham, allows learners to maintain their attention on the world around them while their device is presenting appropriate content, options, and resources that support their learning activities.

For example, learners following a particular course in an art museum will see different content and options being presented to them as they move around the galleries and exhibits. The context awareness subsystem will use contextual information such as location, time, and learner profiling to make recommendations to the content delivery engine about what items should be displayed. Services can also be recommended directly to the user interface: a student who has been struggling with a particular question for some time will be presented with the option to start a chat session with another learner, who may be someone from their own study group, another visitor to the gallery, or perhaps an online student who is visiting the gallery remotely.

Our activities in the MOBIlearn project are centred on specific learning scenarios, of which the art gallery scenario is one example. We have found it useful to describe an underlying model of context that has informed our architecture and enabled relevant discussions between project partners about the use of contextual information in the system as a whole.

3.1 Model of Context

For MOBIlearn, the purpose of context awareness is to enable learning on mobile devices, and so our approach to describing context and applying this description to producing a usable software architecture is based on this focus. Figure 1 shows the basic hierarchy for our description of context.

Instead of a rigid definition, our intention is to provide a hierarchical description of context as a *dynamic process with historical dependencies*. By this we mean that context is a set of changing relationships that may be shaped by the history of those relationships. For example, a learner visiting a museum for the second time could have his or her content recommendations influenced by their activities on a previous visit.

A snapshot of a particular point in the ongoing context process can be captured in a *context state*. A context state contains all the elements currently present within the ongoing context process that are relevant to a particular learning focus, such as the learner's current *project*, *episode*, or *activity* (see [7]). A learner may at any one time be engaged in a number of simultaneous activities and episodes that relate to one project, and they may have several ongoing projects each of which has its own set of relevant activities and episodes. It is therefore important, from a design perspective, to clearly identify the focus for our current implementation of context awareness.

A *context substate* is the set of those elements from the context state that are directly relevant to the current learning and application focus, that is to say those things that are useful and usable for the current learning system.

Context features are the individual, atomic elements found within a context substate and each refers to one specific item of information about the learner or their setting (for example current learning task or location). In our description of context, context features an indivisible and refer to only one item of relevant information about the learner or their setting.

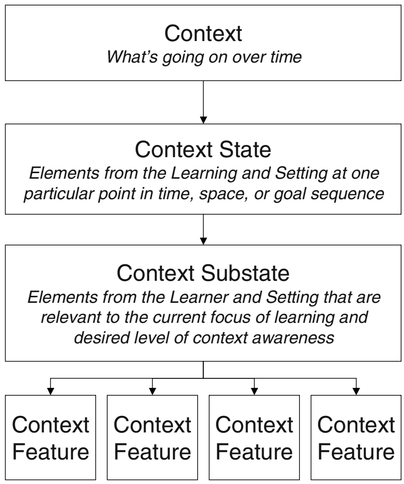


Fig. 1. Context hierarchy

Note that so far we have not specified what elements of the learner's current context we are interested in – this is done on a scenario by scenario basis to allow for maximum flexibility and to better match the context awareness to the learner's needs.

Contextual information is also made available to other parts of the MOBIlearn system by means of XML (eXtensible Mark-up Language) documents in an agreed format. At any given time, the current context state is represented as a nested set of context features, all described in XML form. An XML schema for this XML object is an agreed format that allows all components of the MOBIlearn architecture to access this information as and when it is required. Storage of a set of timestamped XML context objects provides the historical context trace that can be inspected and used by subsequent sessions.

3.2 Context-Awareness Architecture

Figure 2 provides a basic illustration of how the MOBIlearn context awareness subsystem relates to other architecture components and how it provides recommendations to the user. A learner with a mobile device is connected to a content delivery subsystem, which in turn is linked to the context engine. The context awareness subsystem (CAS) collates contextual metadata from sensors, user input, and a user profile. A set of software objects then use this metadata to perform

evaluations of the metadata available on a set of learning objects, options, and services. These evaluations lead to recommendations that are then used by the delivery subsystem in determining which content to deliver to the learner. Note that user input to the system is acknowledged as an input source of contextual data: meaningful context is difficult to establish and we aim to include the learner themselves in the context gathering process.

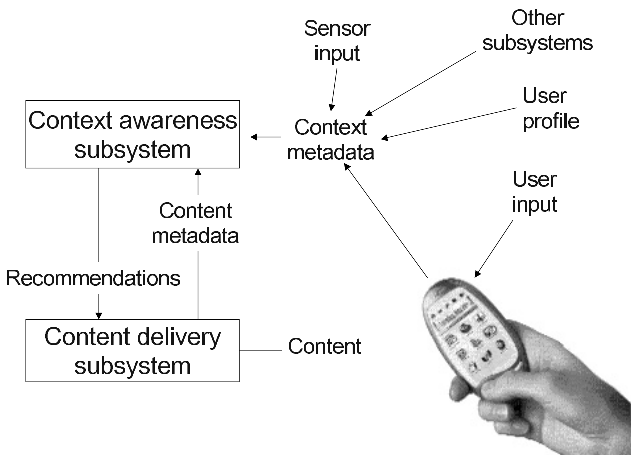


Fig. 2. Context awareness in action

The basic cycle of operation of our context-awareness system is as follows:

1. *gathering* and *input* – of context metadata
2. *construction* – of context substate
3. *exclusion* – of unsuitable content
4. *ranking* – of remaining content
5. *output* – of ranked list of content.

The CAS comprises a set of software objects called context feature objects (CFOs) that correspond to real-world context features relating to the learner's setting, activity, device capabilities and so on to derive a context substate, as described above. Data can be acquired through either automated means (for example sensors or other software subsystems) or can be input directly by the user. This context substate is used to perform first exclusion of any unsuitable content (for example high-resolution web pages that cannot be displayed on a PDA) and then ranking of the remaining content to determine the best *n* options. This ranked set of options is then output to the content delivery subsystem.

3.3 Types of Context Features

Context feature objects are either excluders or rankers. Items of content that are deemed entirely inappropriate for the current context are excluded. That is to say they are removed from the list of recommended content and not subject to any further

consideration. Content remaining in the list after the exclusion process is then ranked according to how well it matches the current context. The ranking process simply increments the score of each item of content that has metadata matching the stimulus values of any particular context feature. The size of the increment depends on the *salience value* of the context feature doing the ranking. Individual CFOs can have their salience values changed so that they exert more influence on the ranking process. Any individual CFO can be de-activated at any time so that it has no effect on the exclusion or ranking processes.

A CFO has a set of possible values, and an indicator of which value is currently selected. It is also possible for CFOs to have multiple sets of possible values, with the current active set being determined by the current value of another linked context feature. Whilst this has no bearing on the recommendation process, it is important in terms of providing an inspectable model of the context state to the user, who can observe the influence of one context feature on another. For example, options relating to current activity can change depending on the user's current location.

3.4 Linked Context Features

Each context feature object responds to only one metadata tag and performs either an exclusion or ranking function. To achieve more complex filtering of content, CFOs can be linked together so that their function can depend on the state of other context feature objects. For example, we might choose to have a context feature object that excludes content based on its file-size – such CFO should be active if the learner is using a low-bandwidth connection, but should remain quiescent if a high bandwidth connection is available. By creating a context feature that responds to bandwidth availability and allowing it to control the status of the context feature that responds to file-size, we can easily create a pair of context features that respond to a more complex context. This linking process is transparent to the user and to individual CFOs, so long chains can easily be created to cope with complex situations.

3.5 Output

The ordered list of ranked items of content is passed to delivery subsystems for use in determining exactly what content should be made available to the user. In this way, the context-awareness sub-system has no way of specifying exactly what is made available – the system is intended only to make recommendations to the system and to the user. This method of recommendation is preferred so that should the system make a mistake, and make inappropriate recommendations, its output does not override selections made elsewhere in the system (for example, the user might specify a particular page of content and then not want that item to be replaced by another).

3.6 Metadata Schema

We have developed a metadata schema to facilitate the appropriate storage and transfer of contextual data among the different components in the MOBIlearn system. This schema maps on to our hierarchical description of context itself and offers a

generic and reusable template for exchanging data about the current context. This schema is also intended to map very closely onto the underlying design of our current software architecture – all context feature objects in the system are implemented as Java objects with attributes that mirror those shown in the schema. Translating from Java object attributes to XML is therefore an efficient way for the system to make its current state available to other system components. A diagrammatic representation of this context schema is shown in Figure 3.

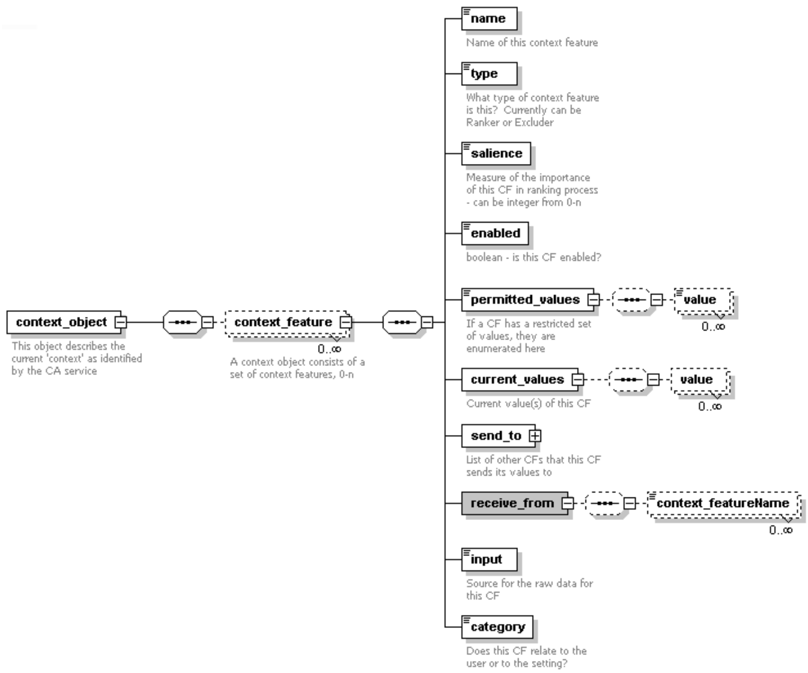


Fig. 3. Context meta-data schema

The root element, ContextObject, is the entire set of all of the contextual features that are currently maintained by the system. Each context feature element corresponds to a software object that listens for changes in a specified feature of the real-world context and responds accordingly. Typically this response will be a re-ranking of the available content to match the new context. This schema is deliberately designed so as to not be prescriptive in itself about which elements of the context state we are currently interested in. Each context feature element contains sub-elements that allow the description of each software context feature in terms of its name, type, enabled status, current and permitted value(s), salience value, input source, and category, as well as the set of other context features that this feature can send to or receive from. The element ‘category’ is used to indicate whether this feature relates to environmental or user data – we have identified both of these sources as important for enabling context aware learning applications.

We address the need to monitor and respond to context over time by storing a series of context objects, each of which has its own timestamp and can be marked

with any other data that relates it to a particular episode, activity, or task. Our aim is to use these 'context traces', made up of groups of context objects, to influence the context of a future use of the system. For example, a learner who has already visited an art gallery on a previous occasion would be able to retrieve their previous context trace and use it to better guide the system for this visit. The previous context would become part of the current context, thus satisfying our identified need for historical dependencies.

With contextual metadata available in this XML format, it is a relatively easy process to apply the exclusion and ranking process outlined earlier. Metadata relating to available learning objects is read into the system as a series of XML documents adhering to the IMS 1.2 schema for learning object metadata [5], and comparison of these two sources of metadata yields contextually relevant recommendations. As we have already found, metadata relevant to *mobile* learning are not fully addressed by the IMS schema, and so we worked on extending this and other schemas to rectify this problem. For more details of this work, see [2].

4 User Trials

We have run some small scale user trials to assess the impact of context aware content delivery on users' experiences and provide some formative evaluation of our work so far. The results of these trials will inform the design of our next prototype.

4.1 Software Setup

All participants used a prototype of the context awareness subsystem implemented in Java. The prototype comprised a single server or management application connected to several clients.

The manager application allowed the experimenter to manage and monitor several participants simultaneously, updating their location and observing the current question they were working on as well as previous completed questions. This application ran on a laptop PC that was used by the experimenter during the session. The laptop was connected to the client software over the wireless network using socket communications.

The CAS client ran on the tablet PCs (Fujitsu Stylistics) that were given to the participants for the session. This application ran beside the Internet Explorer browser and offered recommendations of content, questions, and people that were deemed relevant to the learner's current context. Context was determined from a combination of location of user, location of other users, current question being answered, and previous questions answered.

The system provided recommendations of content, questions, and communication with other learners, depending on the participant's current location and question. For example, a learner standing in front of La Primavera would see content relevant to that painting near the top of their content list, with the top item being most relevant to the La Primavera *and* their current question. If another participant who had already answered the current question was also at La Primavera, the system would suggest talking to them.

4.2 Method

Participants were divided into groups of 2-4 and asked to play the role of art history students following a study guide in an art gallery. They were each given a Fujitsu Stylistic tablet PC and were told that they would get help in finding the answers to the questions from the context awareness system running on the tablet. The basic functionality of the system was explained to them and they were given a brief demonstration of how to use it. Participants were asked to move around the simulated art gallery containing 6 paintings whilst trying to find the answers to 8 questions given to them at the beginning of the session. Participants walked around a set of 6 paintings located in a small room intended to represent an art gallery. They were told that their location relative to specific paintings would affect the recommendations given to them by the system.

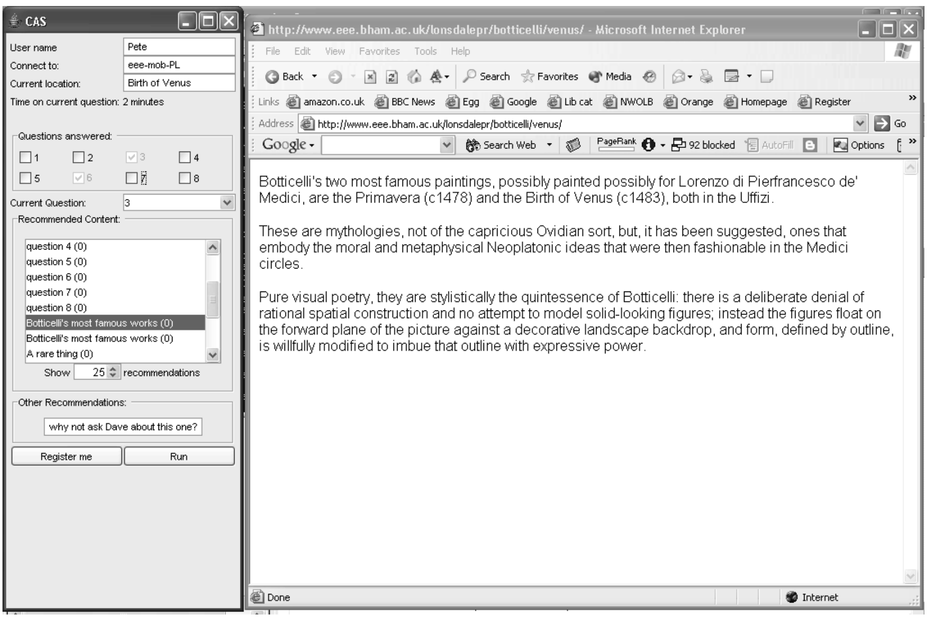


Fig. 4. Context awareness client

During the session the experimenter employed a "Wizard of Oz" evaluation method, monitoring each participant's location and updating their client software using a remote management application. The experimenter was also able to monitor which question each participant was currently working on and which questions had already been answered. At the end of the session, the concept of context aware content delivery was discussed with the participants and they were asked for feedback about their use of the system in an informal, free-form interview.

4.3 Results

Feedback was gathered from users about the usefulness and usability of the system. This feedback was used to derive a formative evaluation of our current implementation. The following issues were identified:

- *It worked*: Most users were able to quickly find relevant information and successfully answer the questions.
- *Interface and representations:* Many users were confused about what we were trying to represent with the interface, and were not sure why their recommendations were changing or how they could best use the recommendations list to answer the questions.
- *Understanding*: some people weren't quite sure why the system did what it did, and were surprised by the constantly changing list of options. Demonstration and explanation did not seem to help with this – when there was a misunderstanding it was due to a lack of intuitiveness about the display of the context-dependent recommendations
- *Distraction vs Engagement*: offering multiple choices either led to sidetracking or encouraged people to further their exploration of the content. Both of these suggest that users were engaging with the experience, but this could become a concern if we are trying to design a specific programme of learning. Options that distract users from their current task focus need to be avoided, and so it is possible that some limits need to be set on exactly how much contextually based recommendation is done.
- *Mixed content:* there is a need to distinguish questions, content, physical resources. Offering recommendations of all of these in a single, integrated display seemed to be confusing, especially in combination with the lack of an intuitive, easily grasped model of what was actually going on and why.
- *Temporal context*: Context is often used in a snapshot sense: what is happening now, where am I at this moment, and so on. However, there are many much longer-term aspects to context (e.g. task, learning progress, life goals) and it is not clear how to best represent and use this information in the context system. The fundamental issue is that we need to be able to model and then provide support for users across multiple activities, episodes and projects, with the history of previous support playing an integral role in determining future actions.

5 Conclusions and Next Steps

The context awareness system that we have developed demonstrates that it is useful and works so that people are supported in their actions. However, it is clear that there is not a sufficiently effective model of context communicated to the user, so that they are often confused as to why the system is changing its recommendations to them. This is partly an interface issue, where the separation of the different parts of the system is unclear, and partly a conceptual one, since users are not used to systems being dynamically adaptive on such a scale. It may be that hiding more of the

workings of the context engine and simply presenting the results using an appropriate metaphor would be more effective; for example, using an avatar gallery guide, which users could easily ascribe some form of 'intelligence' and hence more easily accept changing suggestions.

We are now focussing on extending the context engine and the sensor inputs, and integrating them with the meta-data schema and set of learning objects, to provide a rich environment for more extensive user trials and evaluations. We will also continue our related work on developing new knowledge-based approaches to implicit modelling, to provide more effective models without overloading users with questions. The intention is to allow the system environment to develop models without overtly intruding on its users. Knowledge-based systems inevitably make mistakes, and we are exploring how to resolve conflicts between parts of the system that reach different conclusions and how to cope with interpreting and reconciling heterogeneous sources of data.

References

1. Bormida, G. D., Lefrere, P., Vaccaro, R., and Sharples, M. (2002). The MOBILearn Project: Exploring New Ways to Use Mobile Environments and Devices to Meet the Needs of Learners, Working by Themselves and With Others. In *Proceedings of European Workshop on Mobile and Contextual Learning*, Birmingham, UK, The University of Birmingham, UK.
2. Chan, T., Sharples, M., Vavoula, G., and Lonsdale, P. (2003). Educational metadata for mobile learning. In *Proceedings of Second IEEE Workshop on Wireless and Mobile Technologies in Education*.
3. Chen, G. and Kotz, D. (2000). A survey of context-aware mobile computing research: *Technical Report TR2000-381*, Dartmouth College, Hanover, NH
4. Dourish, P. (2004). What we talk about when we talk about context. *Personal and Ubiquitous Computing*, **8**(1): 19-30.
5. IEEE 1484.12.1-2002 (2002). Draft standard for learning objects metadata, IEEE 1484.12.1-2002.
6. Lonsdale, P., Baber, C., Sharples, M., and Arvanitis, T. (2003). A context awareness architecture for facilitating mobile learning. In *Proceedings of MLEARN 2003*, London, LSDA.
7. Vavoula, G. and Sharples, M. (2002). KLeOS: A personal, mobile, knowledge and learning organisation system. In *Proceedings of IEEE International Workshop on Wireless and Mobile Technologies in Education*, IEEE Computer Society Press.

Usability of MobiVR Concept: Towards Large Virtual Touch Screen for Mobile Devices

Tiiu Koskela and Inka Vilpola

Tampere University of Technology, Institute of Software Systems, P.O. Box 553,
FIN-33101 Tampere, Finland
{Tiiu.Koskela, Inka.Vilpola}@cs.tut.fi

Abstract. The size of the display is one of the major obstacles to fluent information presentation and management on current mobile devices. The user may not be able to perform basic data handling tasks, such as to find and to compare information, in an efficient and satisfying way. This paper presents a usability study of a novel user interface concept – the MobiVR – and its early prototype with a large virtual display and a finger pointing input method. The purpose of the usability study was to find target applications of the MobiVR concept and the features that need to be developed before the advantages of the concept can be fully utilized. The research was conducted with focus group sessions and usability tests. The study showed the MobiVR concept to suit well for spontaneous information retrieval, for example using mobile Internet. The results will direct the further development of the prototype towards a "virtual touch screen".

1 Introduction

Personal mobile devices with features enabling for example communication, time scheduling, and data management, are already part of the everyday life. The tendency of the development is to provide users with even more means for handling information in a mobile context. The emerging technologies create the concept of continuous presence of and connection to selected people and online services [15]. Thus there is a great need for new interaction solutions via which the services can be provided to the user in an easy and pleasant way. Current mobile devices with limitingly small displays and restricted input methods cannot provide optimal user experience with the new mobile application areas such as mobile Internet browsing.

Interaction can be evaluated among others via the elements of *information presentation* and *management.* The user requirements for information presentation and management rely on theories of human cognitive processes (e.g. thinking, problem solving and memory processes). According to them information should be presented in a suitable way for the user to categorize and compare data, and to make conclusions and overviews. Thus the user should be provided with both detail and contextual information [14], which is a very challenging task considering the small size of mo-

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 252–263, 2004.

bile devices. In addition there is need for efficient and easy ways to manage the presented information, e.g. natural and flexible navigation and control possibilities. The direct object manipulation UI and the WIMP (Window, Icon, Menu, Pointer) UI [9] have been found to be useful with PC's, but again the current interaction methods of mobile devices do not fully support them.

In this paper, we describe usability research of a MobiVR concept [11]. The MobiVR concept presents a novel interaction method, which may enhance the possibilities to present and manage information on mobile devices. The MobiVR concept includes a near eye microdisplay and a tracking system. The microdisplay provides the user a desktop-sized virtual display with average PC-level resolution. The tracking system enables pointing, for example with a finger, as an input method. Thus the MobiVR concept is a means for integrating a large "virtual touch screen" in a mobile device. The purpose of the usability research was to find suitable application areas for the concept and to set the direction for further development via examining acceptability and usability issues of the implemented prototype.

First we give a short introduction to current achievements in information presentation and management in Section 2. In Section 3 we present the MobiVR concept and a description of the usability research. The results are presented in Section 4. ideas for further development are visioned in Section 5. Finally, Section 6 concludes the paper by discussing the key issues and further directions of the research.

2 Previous Research

The main challenge of information presentation is to provide the user with an overview and the context of the detailed information. Information visualization proposes scalable interfaces, such as the ones provided by the fisheye techniques [14]. The magnification effect enables the user to make an overview of the information structure and connections, but as a disadvantage it makes focus-targeting difficult because objects appear to move as the focus point approaches them [1]. Context-aware solutions, on the other hand, aim to filter irrelevant information and show just the area of information that the user needs or is interested of. The detail information can be shown with context, because the focus area is restricted [10]. But due to the restrictions there may be doubts whether user's actions and possibilities are also restricted.

Despite that visual information dominates in the current world, some efforts have been made for solutions for other modalities, such as sonically enhanced interface. In enhancement studies the sounds are used as metaphors and earcons [7], for example. Additional modalities are useful for providing feedback and thus diminishing the visual attention needed for basic control, but they are not a solution for true information retrieval. In addition, the problem in the usage of sound as a part of a user interface is that it makes the tasks and actions public, which can invade the personal nature of mobile devices. Tactile information can be used for somewhat the same pur-

poses as sound, for example in notification, monitoring the state of a device, and for certain level of feedback [8]. The advantage of tactile information is that it remains private.

Virtual reality has tackled the information presentation problems rather successfully broadening the display area without affecting the size of the device. The solutions are commonly in a development phase, and the design challenges are related more to information management issues, like how the virtual display can be reached and controlled via the mobile device. One approach is *outside-in systems*, of which a representative is the Peephole-Display concept designed by Ka-Ping Yee (2003). In Yee's design a small display of a mobile device acts as a window, via which a large virtual workspace can be looked at one part at a time. The main advantage of the concept is that it provides flexible information management by enabling fluent navigation via gestures (moving the mobile device, that is the peephole, on top of the virtual workspace). The disadvantage is that the user can reach and manage only the part of the information that is currently seen in the peephole [16].

Virtual reality aims for more effective information presentation and management by immersing the user in a virtual environment via wearable computing, such as data gloves and head mounted displays (HMDs) [5]. Wearable computing tries to enable the utilization of a computing system simultaneously with other activities. Nevertheless multitasking is rather challenging, because in many cases the interface seems to block the user's view [4]. Also, physical and social issues affect the acceptability of a wearable computer; for example HMD has been found to cause simulation sickness during the performance of some tasks (for example stereoscopic game playing) [3], and today's prototypes, for example data gloves, can provide the image of a cyborg [6].

Therefore, even though the trend is towards integration and invisibility of computing, today's everyday mobile computing is still relying on handheld devices. In practice, virtual reality can provide a solution to the display size problem, but the input methods remain rather complex and difficult for mobile information management.

3 The MobiVR Concept

This study focuses on the MobiVR concept developed in the Institute of Signal Processing at Tampere University of Technology. The concept provides the user with a full-size virtual display on a mobile handheld device. The user's finger can act as pointing device (Figure 1). The MobiVR concept is based on a wireless communication channel, a handheld (not head mounted), a near-eye microdisplay, and a tracking system to track the selected pointing device (e.g. the user's finger) [11]. The MobiVR concept is designed for a *handheld* device, in order to gain social acceptability by supporting similar usage patterns and situations as with current mobile devices (i.e. mobile phone).

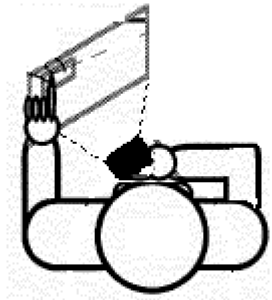


Fig. 1. The user's finger can act as a pointing device on the virtual touch screen.

The MobiVR concept aims to provide a GUI-based full-resolution virtual screen in a handheld device. The handheld device provides a private virtual screen, when it is held in front of the eye. The large virtual display enables the presentation of a large amount of information and thus provides means for overview of both contextual and detail information. The large virtual display makes it also possible to use GUIs and high resolution, with which users are accustomed in a PC environment. In addition, the MobiVR concept enables adjusted direct object manipulation.

The MobiVR concept was implemented into an early prototype with modified low-cost components that are already available in the end-user market (Figure 2). The technical details can be found in Rakkolainen's paper [11].

Fig. 2. The early prototype of MobiVR.

The characteristics of the early prototype are a two-eye microdisplay and wire connection to the computer's display adapter. The input method is divided for two hands; one for pointing and the other for making the selection. The pointing finger needs a separate IR reflector that the camera, which is integrated to the prototype, can track as a cursor. The same hand that holds the device does the selection. The selection button is placed in the top centre of the device. The size of the prototype is 145x85x50 mm and it weighs 140g. The microdisplay and tracking camera technologies are covered with a plastic coat.

4 Usability Research of the MobiVR Concept and Prototype

This study is the first usability research for the MobiVR and it was carried out during September to December 2003 at the Institute of Software Systems of Tampere University of Technology. The development of the MobiVR concept and prototype had so far based only on the innovation and creativity of the designers. Thus the primary focus of the research was to provide information for the development process by defining the acceptability of the MobiVR concept and its suitable application areas. This research should be considered as a case study and, in order to reach statistical validity, additional user research should be done with a larger user sample.

The specified objectives of this study were the following:

1. *Acceptability and utility of the MobiVR concept:* What are the user's needs for information presentation and management in mobile computing? What are the potential application areas for the MobiVR concept?
2. *User experience of the MobiVR prototype:* How efficient and useful seem the information presentation possibilities to be for the user, and how easy is information management via the prototype? How pleasant is the usage?

The research process consisted of two phases relating to the two main objectives. *In the definition phase,* focus group sessions were held to gather information about the users' needs for information presentation and management in mobile computing, and whether the MobiVR concept can respond to those needs. *The evaluation phase* consisted of usability tests of the MobiVR prototype. The tests concentrated on evaluating the feeling of control and efficiency provided by the input method, and the value of the large virtual display. In addition, the users' subjective satisfaction was charted.

Three focus groups consisted of *advanced users*, *young users* and *non-technically oriented users* (total of 6+6+3 Finns). The advanced users had previous experience on using either WAP-services or a camera implemented on a mobile phone. Their mobile phone usage was daily for both work and pleasure. The group's profile was formed in order to reach an opinion of people that are most likely among the first active, everyday users of new mobile technologies. Young users were 15-16 year old high-school students, who had seen the new mobile phone features (such as camera) used, but not actually used themselves. They were selected to give information about the possible trends and application areas among teenagers. The non-technically oriented group consisted of people that did not follow the technical development of mobile phones closely and used only the primary features (voice call and SMS) of their mobile phones. They were selected to give information on whether the MobiVR concept would be easy to approach or accept for everyone.

Two researchers, one of whom acted as a moderator as the other took notes, held the focus group sessions. Each session took two hours and they were also videotaped. The sessions followed two steps. In the first step the participants were asked to tell

about their current usage of mobile computing (phones, PDA's etc.). The moderator encouraged the participants to discuss the usage contexts, situations, purposes and especially the problems the participants had accounted in the usage of mobile devices. The goal for the discussions was to define general requirements for the mobile computing.

With the second step the MobiVR concept (a near eye micro display, virtual display and tracked pointing input method) was introduced to the participants via two scenarios of mobile device usage, where the MobiVR concept was integrated into a mobile phone. In the first scenario still images presented a MobiVR usage situation in which the user browsed online timetables for buses while waiting at the bus stop. The second scenario was a short video about two colleagues meeting in a café and seeking information on a forthcoming event (Figure 3).

Fig. 3. A scenario "Seeking information on-line using the MobiVR concept integrated to a mobile phone."

After the scenarios the users were asked about the first impressions and the acceptability of the MobiVR concept. The moderator led the participants to discuss also about how the MobiVR concept would answer to the usability problems mentioned in the earlier mobile computing discussion.

In *the evaluation phase,* the usability tests were held in a laboratory as single user tests with 8 test users (Finns). Users were mainly advanced users (6), because the tests focused on the actual usage not on the learning phase. Learning phase may take longer with novice users. The tests were performed using the Think Aloud protocol, which experts state to present 80% of the usability problems with 5-6 test users [9]. Similarly to the focus groups, two researches, a moderator and the other taking notes, carried out the tests and the test sessions were videotaped. One test session lasted for an hour consisting of 5 different tasks, which were followed by an interview concerning the test users' opinions, comments and ideas.

The test tasks were planned to create a versatile user experience. Instead of a specific type of applications, the users' performance was observed focusing on the users' ability to control the prototype, the success to point and select an object and the intuitivism of the information search (data handling). The tasks consisted of the following:

- General information search (Web site of a local sports event): purpose to relax the user and give her/him a time to get used to the MobiVR prototype.
- Specific information search with making conclusions of the material (Web site of a local sports event): purpose to examine a case where the user is confronted with new information in an unfamiliar context (with unfamiliar surroundings and hierarchic structure), and to discover how the MobiVR supports comparing information in order to make conclusions.
- Information search (local bus timetable): purpose to examine how MobiVR suits for cases where the user is already familiar with the application and the information and its structure is somewhat predictable.
- Object manipulation (calculator application): purpose to study the efficiency and accuracy of the object manipulation (virtual keyboard) possibilities of the MobiVR.
- Game for hang-fly simulation: purpose to gather information on the user's experience of the level of the immersion provided by the microdisplay and of the flow-experience in the input possibilities.

5 Results of the Usability Research

The utility and acceptability of the MobiVR concept was charted via focus group sessions and the user experience of the MobiVR prototype was analysed via user tests. The results are presented according to user reactions and actions. The tasks in the user tests were only to simulate different use cases and are not reported individually in the results.

5.1 Results on Utility and Acceptability of the MobiVR Concept

The focus group participants wanted the mobile devices to provide them means for *communication*, *control of everyday life*, and *information management*. The participants emphasized the possibilities to communicate by messaging and stated that, for example, pictures are much easier and more effective for presenting a feeling or an experience than plain text. Thus it can be said that the communication devices are desired to be expressive tools with the capacity to present visual information. The control of everyday life includes managing connection, time, money and in particular information. Knowledge management - information search, retrieval, organizing and storing – is needed both for work and leisure. This raises requirements also for effective input methods even while the user is in a mobile context. More commonly the

source for information gathering is the Internet, even when the non-technically oriented users are concerned.

The participants looked surprised to see a mobile phone raised in front of the eye in the scenarios. So far not even camera phones have required such an action. This new interaction with a mobile device raised concerns of its *social acceptability*. The participants were afraid that a near eye display would immerse the user in the virtual world and cause isolation from the world outside, even though the users had only one eye covered in the presented scenarios. The subjects saw the isolation also as a small threat for security, because the user is unable to fully observe what is happening in the surrounding environment. The participants considered also the input method, pointing in thin air, to be odd. Then again, the users saw the analogue to the hands-free paradigm, which had also been somewhat embarrassing with mobile phones in the beginning, but is now rather acceptable.

The second area for the participant's concerns, right after social issues, was *health issues*. The subjects wondered whether the near eye display would affect the user's sight with occasional or permanent damage.

According to the participants the MobiVR concept would suit best for *short-term ad-hoc information search* like seeking information about bus timetables or the currently presented movies at the cinema nearby. Also applications containing *context aware information*, like maps and location-based guidance, were seen as a potential. Surprisingly nobody suggested games.

Even though suitable and rather appealing application areas could be found for MobiVR, the acceptance cannot be assured in general. Non-technically oriented users argued intensively about utilizing MobiVR in real life. For every suggestion, for example using MobiVR to get recipes from the Internet, there was a counterproposal, like why cannot we use books instead. Some users in the group considered that there is no need for MobiVR (or any new solutions for mobile devices) and all the ideas feel artificial. Then again the advanced users were enthusiastic about the concept and suggested several further development ideas.

5.2 Results on the Usability of the MobiVR Prototype

The usage of the MobiVR prototype requires *two-handed co-operated coordination.* One hand must support the display device and push the selection button as the other hand is used for moving the cursor. Movement of the display device has also an effect on the cursor position, because the tracking system is inside the device. This caused a problem, because the movement is almost inevitable as the user pushes the select button on top of the device. Thus the users lost the focus of the cursor. The users stated that it would be more intuitive for them to do both pointing and selecting with only one hand.

As the display is virtual, the users had difficulties in determining the area within which the pointing hand should move. The prototype did not give users *feedback of the virtual workspace*, for example the system did not indicate to the user whether it could track the pointing finger or not. In addition, as the prototype covered both eyes, the users lost the sense of physical distance and position of his own finger. These aspects caused the users to loose the control of the cursor, which was seen as unpleasant delays in the movement of the cursor. After loosing the control the users were forced to calibrate the tracking system (via taking the cursor from top of the screen to the bottom, and from the left edge of the screen to the right), which diminished the efficiency of the usage.

In most of the usage situations the users had to support hands up in the air. This caused *physical strain*, which diminished the value of the large virtual display. The users stated that because of the physical effort they had hard time to focus on the text they were reading. The users stated also that the usage of the micro display required the eyes to adapt, which they found a bit tiresome.

In addition, the users found it inconvenient that the MobiVR prototype did not support suspending a task. The information of the current state (especially cursor) was lost immediately when the user moved the device. Therefore the user could not lower the microdisplay to relax or for example to follow the events of the surrounding world. For some users it was natural to lower the microdisplay every time they discussed with the test moderator. Due to this they always needed to calibrate the cursor and start the task over. The current prototype does not take into account the *social demands* for the mobile devices, which are commonly used in an active environment where the users' attention and reaction are required. Even more, there can be problems to accomplish task-flows with a sequence of steps in a mobile context with multiple interruptions.

Despite the disadvantages of the MobiVR prototype, the users found the size of the virtual display and the goal for a natural input method to be compelling compared to those available in traditional mobile devices, like mobile phones. It took the users for half an hour to get used to using the MobiVR prototype. The users felt that the accuracy and efficiency were improved as they tried different positions in using the prototype. The tasks were more likely to be completed and the actual need for calibrating decreased according to accuracy improvements.

6 Ideas for Further Development of the MobiVR Prototype

The social issues that arose were caused by the fact that the prototype covered both eyes, isolating the user from the outside world. The MobiVR concept would be more easily accepted if it would be integrated into a prototype that *covers only one of the user's eyes* and therefore would allow easier contact with the outside world. Providing means to suspend a task would be an additional way to support contact with the

real world. The prototype could provide the user a way to inform when the task is interrupted and again when it is resumed.

The usage of the current MobiVR prototype demands two-handed co-operated coordination, which the users found to be very difficult to manage. The development of the prototype should be towards supporting true *direct object manipulation*, like pushing buttons and clicking links, with typical easy and intuitive touch screen manners [2]. However, the text input method needs to be considered carefully, because writing with a virtual keyboard and Graffiti is slower and more error prone than with a mechanical keyboard [12]. In addition, information management could consist of *natural gestures*. For example scrolling could be performed via pulling the view down. Focusing of the cursor could be made easier with area cursors and sticky icons. The idea behind area cursors is to increase the size of the cursor hot spot, and as sticky icons are used the cursor movement is reduced when it comes near to an object [1].

The lack of feedback from the virtual workspace made it hard to the users to determine the pointing area and the state of tracking. Feedback from the virtual workspace could be provided via *augmented reality solutions,* in which the display would be partly transparent. The transparent view would enable the user to see the cursor hand and get visual feedback of its position [13]. Another solution would be a *sound-enhanced interface,* in which limited tactile feedback would be compensated by sound [7]. In addition, there could be *visual feedback* such as changes in the cursor colour.

The usage of the current prototype causes physical strain especially when the users are forced to raise their hands over the level of their hearts. The easiest way to diminish the strain is to build a prototype which is used *vertically* as in the scenario presented to the focus groups [Figure 3]. In addition, the researchers' experience is that the way to use the prototype evolves through time and the usage can be rather light after the user is familiar with the possible usage positions. For example, MobiVR can be used facing the virtual display forward and down when the pointing finger can be propped on the table. The input method inspired also the test users to consider other *natural input methods such as speech or eye movement recognition*. The subjects' purpose for these innovations would be to free the other hand from the pointing task. Also wearable solutions could free the user's hands for other tasks.

7 Conclusions

The needs of the focus group participants for mobile computing focused on three application areas - *communication, control of everyday life*, and *knowledge management* - relating mostly to the functions already available on mobile devices. The MobiVR concept seemed to reflect these needs via providing new means especially for information presentation. The subjects saw the concept to have potential, but they

also brought up some acceptability issues concerning mainly the *social acceptability* and the possible *health effects* of the MobiVR.

The same curious and positive attitude towards the MobiVR concept was found also in the usability tests of the MobiVR prototype, even tough the test users found many possibilities for further development. The main usability problems in the efficiency and accuracy of the input method were caused by the demand for *two-handed co-operated coordination* and the lack of *feedback from the virtual workspace*. The *physical strain* and *social demands* on the other hand diminished the value of the prototype's output method, the large virtual display.

The input methods of the MobiVR prototype could be developed further towards natural gesturing and direct object manipulation UI. The feedback provided by the prototype should also be improved for example by enhancing the interface with sound and visual effects, or by providing an augmented reality solution. The MobiVR concept can be seen as a way to provide the user with a large "virtual touch screen" on a mobile device.

The MobiVR concept offers new possibilities to present information in detail and in context. The amount of information that can be presented is much larger than in the mobile devices, that the users are accustomed with. The MobiVR concept enables information presentation in a way that supports the user's data handling tasks, such as comparison of details. The most potential application area includes spontaneous information search, for example in the wireless Internet. In addition, the MobiVR concept can be seen as a means for developing an expressive communication tool, because of its possibilities to present visual information.

The MobiVR research will be continued with the development of a new more advanced prototype, in which the current usability problems will be taken into account.

References

1. Gutwin, C. 2002.Improving focus targeting in interactive fisheye views. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* p. 267-274, ACM Press.
2. Holzinger, A. 2003. Finger Instead of Mouse: Touch Screen as a Means of Enhancing Universal Access. N. Carbonell, C. Stephanidis (Eds.): User Interfaces for All, LNCS 2615, p. 387-397. Springer-Verlag Beerlin Heidelberg.
3. Häkkinen, J., Vuori, T., and Puhakka, M. 2002. Postural Stability and Sickness symptoms after HMD Use. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*. Vol. 1, p. 147-152, IEEE.
4. Lehikoinen, J., and Suomela, R. 2002. Accessing Context in Wearable Computers. In *Personal and Ubiquitous Computing.* Vol. 6, Issue 1, p. 64-74, Springer-Verlag.
5. Lehikoinen, J., and Röykkee, M. 2003. 1D selection of 2D objects in head-worn displays. In *Personal and Ubiquitous Computing.* Vol. 7, Issue? p. 44-52, Springer-Verlag.

6. Mann, S. 1997. Wearable Computing: A first Step Toward Personal Imaging. In Cybersquare Computer. Vol. 30, No. 2, p. 25-32, IEEE.
7. Pirhonen, A., Brewster, S., and Holguin, C. 2002. Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves.* p. 51-60, ACM Press.
8. Poupyrev, P., Maryama, S., and Rekimoto, J. 2002. Ambient Touch: Designing Tactile Interfaces For Handheld Devices. In *Proceedings of the 15th annual ACM symposium on User interface software and technology.* p. 51-60, ACM Press.
9. Preece, J. 2002. Interaction Design beyond human-computer interaction. John Wiley & Sons, Inc. USA. p. 257, 441.
10. Pyssysalo, T., Repo, T., Turunen, T., Lankila, T., and Röning, J. 2000. Cyphone – Bringing Augmented Reality to Next Generation Mobile Phones. In *Proceedings of DARE 2000 on Designing augmented reality environments*, p. 11-21, ACM Press.
11. Rakkolainen, I. 2003. MobiVR – A Novel User Interface Concept for Mobile Computing. In *the 4th International Workshop on Mobile Computing (IMC' 2003)*, p. 107-112.
12. Roeber, H., Bacus, J., and Tomasi, C. 2003. Typing in Thin Air The Canesta Projection Keyboard – A new Method of Interaction with Electronic Devices. In *CHI ´03 Extended abstracts on Human factors in computing systems, session: Short talks: Interaction Techniques for Handheld Devices,*. p. 712-713, ACM Press, .New york, USA.
13. Sasaki, H., Kuroda, T., Manabe, Y., and Chihara, K. 2000. Augmented Reality Based Input Interface for Wearable Computers. J.-C. Heudin (Ed.):Proceedings of the Virtual Worlds: 2nd International conference, Paris, France, July 2000, Springer-Verlag Heidelberg.
14. Spence, R. 2001. Information Visualization. ACM Press. P.112 and 117.
15. Väänänen-Vainio-Mattila, K. 2002. The future of mobile communities: Evolution towards continuous presence. M. Jones (Ed.):ACM SIGCHI Bulletin – a supplement to interactions, vol 2002, May-June 2002, p.9. ACM Press, New York, USA.
16. Yee, K-P.2003. Peephole displays: pen interaction on spatially aware handheld computers. In *Proceedings of the conference on Human factors in computing systems.* Demonstration session: Interaction techniques for handheld devices p. 1-8, ACM Press.

Handy: A New Interaction Device for Vehicular Information Systems

Gennaro Costagliola[1], Sergio Di Martino[1], Filomena Ferrucci[1], Giuseppe Oliviero[2], Umberto Montemurro[2], and Alessandro Paliotti[2]

[1] Dip. di Matematica e Informatica, Università degli Studi di Salerno, Italy
{gcostagliola | sdimartino | fferrucci}@unisa.it
[2] Elasis S.C.p.A. - Veicolo SEE, Italy
{giuseppe.oliviero | umberto.montemurro | alessandro.paliotti}@elasis.fiat.it

Abstract. The design of interfaces for automotive information systems is a critical task. In fact, such design must take into account that user is busy in the primary driving task, and any visual distraction determined by telematics systems can cause serious safety problems. To limit such distraction and enhance safety, in this paper we propose a novel multimodal user interface. The key element of the proposal is a new interaction device, named *Handy*, conceived to exploit the driver's tactile channel to minimize the workload of visual channel. Moreover *Handy* is suitably integrated with the graphical user interface, which is characterized by a reduced number of choices for each state and has been designed in agreement with the self-revealing approach.

1 Introduction

Current vehicular telematics systems are providing even more functionality. Indeed, while earlier models were meant for providing only some route calculation, the most advanced models, such as *GM OnStar*, *Fiat Connect+*, or *BMW iDrive*, allow us to connect to the WWW, check mail, play MP3 or DVD, etc… For these reasons, such systems are often referred as *Intelligent Transportation Systems* (ITSs).

When realizing these systems the design of user interfaces is definitely the main challenging issue. Indeed, the interaction with automotive telematics systems is somehow far to be deeply understood and we cannot rely on neither a standard paradigm nor widely accepted interaction devices.

The experience and the well-established interaction metaphors for traditional desktop environments cannot be transposed in the vehicular domain, where specific issues have to be taken into account. These derive from the fact that end-user is normally busy in the demanding and mission-critical task of driving and the interaction with telematics systems holds in concurrency with such a primary task. Moreover, this interaction involves some visual, manual and cognitive resources. So, it is mandatory that it does not require significant visual workload that can distract driver from his/her main activity, with potentially fatal consequences.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 264-275, 2004.

As the road safety should be the most important aspect when developing telematics systems, it is now becoming clear that the next-generation of automotive applications will require substantial usability enhancements. Thus, currently many complementary efforts are committed, including the proposal of numerous standards, recommendations and guidelines that address the safety of in-vehicle telematics systems [e.g. 6]. Such proposals provide some guidance to designers and some hints to limit driver distraction, but many of them are “unverifiable, incomplete and under-specified” [17]. A discussion on some of the most relevant safety standards and guidelines is provided in [17].

Large efforts are also being devoted to the definition of new paradigms and interaction devices. Integrated multimodal interfaces deserve special interest since they are able to exploit different user’s sensorial channels, thus minimizing visual workload [1].

In this context, the Elasis research centre of the Fiat group and the Department of Mathematics and Informatics of the University of Salerno jointly developed a project aimed to define an innovative and user-friendly interface for the next-generation of telematics systems, which should be easy and cost-effective to industrialize in the next two-three years.

In this paper we present the main results of this collaboration. In particular, we describe the innovative interaction device, named “*Handy*”, conceived by keeping in high priority the safety issues. Indeed, in order to minimize as much as possible the visual workload induced by the system, the proposed interaction device has been designed to exploit the users tactile channel. Moreover, we describe the characteristics of the multimodal interface which can exploit the specific features of *Handy*.

The rest of the paper is organized as follows: in the next section we will discuss the main issues of HMI in the automotive field and some significant interaction devices. Then we will present the *Handy* system, and the resulting graphical and tactile interface. Some final remarks and a discussion on future work will conclude the paper.

2 Automotive Human Machine Interaction: A New Research Area

It is widely recognized that most HCI techniques and approaches established for traditional desktop applications, turn out to be inadequate for the automotive domain [1]. This is due to three main factors:

1. User/driver can dedicate only a few burst of his/her attention to interact with telematics system [2]. So, while for desktop applications the UI designers can make the assumption that user is mainly focused on interacting with the system, when dealing with ITSs they cannot rely on a significant user’s attention, because (s)he is mainly concentrated on the primary driving task, which requires a considerable amount of visual and cognitive workload.
2. Automotive displays can show only a reduced amount of information. This is due to essentially two reasons. The former is that these systems have limited output capabilities. Displays usually are between 5” and 7” and have a poor QVGA resolution (320x240 pixels). The latter concerns with the directives (e.g. [6]) and guidelines about the ergonomics of information presentation, which have been

issued by many global institutions to avoid visual overloads, and force to reduce the amount of information to display in any time.
3. Automotive telematics systems, like other ubiquitous computing applications, cannot rely on an input pointing device such as a mouse or a trackball, making the current implementation of the point-and-click paradigm no longer adequate. Instead, some new interaction devices, paradigms, and metaphors are required.

Thus, there is the need of establishing new techniques and approaches which carefully take into account the specific issues of such in-vehicle information systems. This is a very challenging task for UI designers since it is necessary not only to consider the driver interaction with the interface but also to understand the effects of this interaction on driver performances. Indeed, it is worth taking into account that such an interaction holds in concurrency with the driving primary task and involves some visual, manual and cognitive resources [2]. This determines a reduction of the attention devoted to the primary task, with an overall decreasing of the road safety [3]. "Safety" is definitely the specific and most important requirement for the development of in-vehicle interfaces. Nevertheless, it is obvious that it can benefit from some features such as usability, intuitiveness, consistency, and naturalness, but above all requires that the interface does not absorb significant visual and cognitive resources. To face the problem, in the last years, the use of Head-Up Displays (HUD) have been proposed, thanks to the positive feedbacks coming from military applications [16]. Currently many research efforts are being devoted towards the evaluation of the benefits of such devices and BMW and Chevrolet are going to offer HUDs as an optional for their top-class models. The main advantages of automotive HUDs include increased visual attention devoted to the road and reduced reaaccommodation time, particularly for the older drivers. Nevertheless, some studies have proved that symbols shown on HUDs can mask safety-critical targets in the driving scene, and that responses to external targets can be degraded due to the processing of information from a HUD image [16]. Another fundamental drawback is that HUDs are very expensive to industrialize, and thus not suitable for a broader adoption.

To avoid visual workload inducted by ITSs, large efforts are also being devoted to the definition of multimodal interfaces. The main goals of these approaches are to exploit the other user's sensorial channels in order to not affect the driver visual workload [1]. In this way, user can look at the road, and in the meanwhile can interact with the system using the auditory and/or the tactile channels.

To design such multimodal interfaces several presentation issues have to be addressed that directly affect the safety and usability, such as:

- modality (auditory, visual, and tactile),
- format (textual, iconic, tone, voice, etc…),
- time (start time, duration, frequency, etc…) [4].

Moreover, many efforts also are being devoted to design more suitable interaction devices. In the following some relevant automotive controller devices are discussed.

Even if vocal interfaces are becoming more diffuse, the manual controllers are still the primary channel used to interact with the telematics systems. This means that their design is a critical task, because hostile controllers will induct much more distraction in the user, leading to an overall reduction of safety.

In the earlier automotive systems, the interaction was mainly mediated by buttons, rotaries and switches. But soon it becomes clear that newer interaction paradigms and devices were needed to enhance safety and usability of those systems. Currently, automotive manufacturers propose on the market mainly two kinds of devices for interacting with telematics systems: the ones based on touchscreens, and the ones based on knobs and switches.

Though the former approach can seem to be a very natural way to interact with a system, several concerns have been expressed for the in-vehicle use of touchscreens since they require a considerable visual attention in order to locate and select the required inputs and lack of tactile and kinaesthetic feedbacks [7].

In the last years various car manufacturers, such as BMW or Audi, have introduced on the market some novel controllers which are based on knobs and switches. In particular, presently the most innovative approach is the BMW *iDrive*. The heart of this system is a single multipurpose controller that can move forward, backward and sideways and be rotated like a knob and depressed like a button. However, this solution has been strongly criticized for its complexity. For example, a significant article appeared on the New York Times was titled “Driven to Distraction” [8]! One of the main drawbacks of *iDrive* is that in any time user has to select an action among 10 options (up, down, left, right, the four diagonals, rotation and press), thus requiring significant cognitive resources, since the typical short-term capacity [10] is overflown.

3 The Handy Device

The aim of the collaboration between the research centre Elasis, and the University of Salerno was to define an innovative interface allowing drivers to manage the functionality of next-generation telematics systems.

The main requirements for the system were:

- Minimize the distraction (and in particular the visual workload) inducted by the system;
- Easy to use for naïve users;
- Quick to use for expert users;
- Easy and cost-effective to industrialize.

As a result of this collaboration, the *ADvanced –Human Machine Interface* (*AD-HMI*) has been proposed. The main characteristics of the proposal are a novel interaction device, named *Handy*, and a multimodal user interface, encompassing audio, visual and tactile sections. The main rationale behind the solution was to exploit various sensorial channels, specially focusing on the tactile one. Indeed, even if there are no specific studies in the literature about the adoption of haptic control interfaces for ITSs, it is argued that considerable benefits can be gained from making greater use of the tactile channel [7].

In the following we will present the main features of *Handy*, while next section will be devoted to illustrate the corresponding multimodal interface.

As we have noticed in the previous section the point-and-click paradigm is no longer adequate to interact with ITSs. This makes in-vehicle interfaces deeply different from the ones of traditional PC applications, resembling instead the interfaces of cellular phones. Indeed, usually a GUI of ITSs is structured as a sequence of screens (or states), arranged in a hierarchical fashion. Each of them allows us to access specific information and features. Generally such a structure is a multi-rooted tree, where each root corresponds to the starting screen of a module (navigator, audio, etc…) composing the ITS. As a result, the interaction with an ITS involves different subtasks, such as moving within the hierarchy of menus, selecting items among those present in the lists, varying the value of an attribute on a continuous domain, choosing a spot on a map, etc…

To perform any of the above (sub)tasks, user has to accomplish the following actions:

1. (s)he has to search the hardware control suited for the specific task,
2. then (s)he has to place his/her hand or fingers on that (typically small) control, and
3. finally (s)he has to interact with it in order to achieve his/her goal.

This means that each input always encompasses the highly distracting task, where driver takes out the glance from the road and looks at the faceplate of the system, searching for the specific control to handle.

The *Handy* interaction device[1] has been meant to overcome some limitations of current automotive systems. In particular, the main advantage inducted by *Handy* is that user is always aware of both its position and of the displacement of his/her fingers with respect to the interaction controls. Indeed, *Handy* is characterized by an ergonomic, comfortable shape, recalling in some way hand's palm (like the most advanced PC mice) and encompasses a rotary wheel, placed under the forefinger, and four buttons, placed under the other fingers. In Fig. 1 it is shown the left-hand drive version of *Handy*. Obviously, in the nations adopting the right-hand drive, the shape of the device will be the mirror-like of the one depicted below.

Fig. 1. The Handy device, and a detail of the wheel

The natural seat for *Handy* is a bay placed between the front seats or on the driver seat arm rest. So, to grasp it is an activity very similar to grip the gear lever, where driver does not have to allocate visual resources to accomplish the task since (s)he can reach it only relying on the spatial awareness and on the tactile channel.

[1] Handy is currently patent pending.

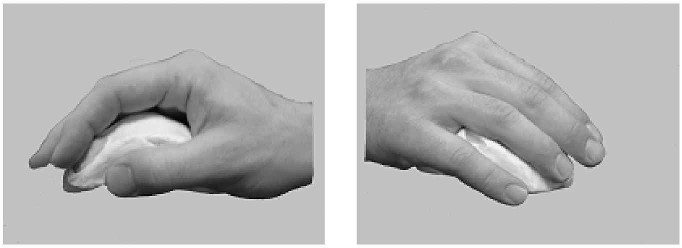


Fig. 2. How to grasp Handy

The others features of the device contribute to significantly reduce the required visual workload. Indeed we have no longer to look for the specific control to handle, since once grasped *Handy* all the controls are suitably positioned under fingers (see Fig. 2).

Special care has been devoted to design the set of controls embedded in *Handy* in order to suitably interact with an ITS. The wheel can be clicked, and has two degrees of freedom, i.e. it can be rotated on the vertical axis and tilted on the horizontal one, like the ones included in the *Microsoft Intellimouse* or in the smart-phone *Sony Ericsson P900*. Moreover, both the wheel and the buttons can return some tactile feedbacks to user, as explained in the sequel.

The rotation and the subsequent click of the scroll wheel can be used to highlight and then select an item among those present in a list, such as a radio station among the ones stored in memory, or a navigator address between the last input. The tilt of the wheel can be used to move the focus among two horizontally adjacent objects, such as a master-detail description of an item. The actions allowed by the wheel agree with the guidelines of vehicular specific HCI [12], as well as the institutional directives on automotive controls [11].

The button placed under the thumb is always used to perform a *Back* or *Escape* functions. For example it permits to come up one level in the hierarchy of menus, or to interrupt an input task. This association is coherent with the occidental stereotype that links the "go back" action with something placed in the lower left direction [5].

The remaining three buttons are mainly suited to permit the navigation within the hierarchies of menus. Their semantic depends on the current state/application and is highlighted in a specific section of the GUI (as described in the next section), and for that reason are named *softbuttons*. The characteristics of *Handy* are summarized in Table 1.

It is worth noting that for each state, *Handy* provides the user with at most 7 actions (horizontal scroll, vertical scroll, scroll click, T1, T2, T3 and TB clicks), thus fitting the 7±2 capacity of the typical user's processing load [10].

Moreover to limit the driver's visual workload, *Handy* exhibits other important features. In particular it exploits the tactile sensory channel of driver, by providing some haptic feedbacks, to communicate information from the system to the user without involving the visual channel. It is recognized that the addition of adequate tactile feedbacks to a user interface can result in many advantages, such as reduced errors, reduced times to complete tasks and lowered workload [15].

In our proposal, haptic feedbacks are used to enhance the user awareness of the system state and to help him/her in the navigation of the menu structure, limiting the visual workload.

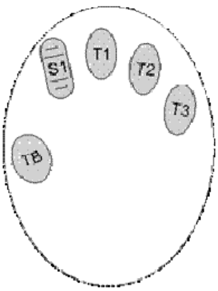


Fig. 3. Naming of Handy's buttons

Table 1. Semantic of the Handy's buttons

Control	Type	Features
T1, T2, T3	SoftButton	The semantic of these buttons depends on the active module and state of the system, and is described by a label placed in a specific section of the GUI.
TB	SoftButton	This button is associated to an "escape" or "abort" function. It allows us to come up one level in a menu hierarchy, as well as to abort a current interaction.
S1	Clickable and tiltable scroll wheel	The semantic associated with the rotations, tilts and clicks of this element depends on the active module and state of the system, and is described by a label placed in a specific section of the GUI.

In particular, the scroll wheel provides the following kinds of tactile output:

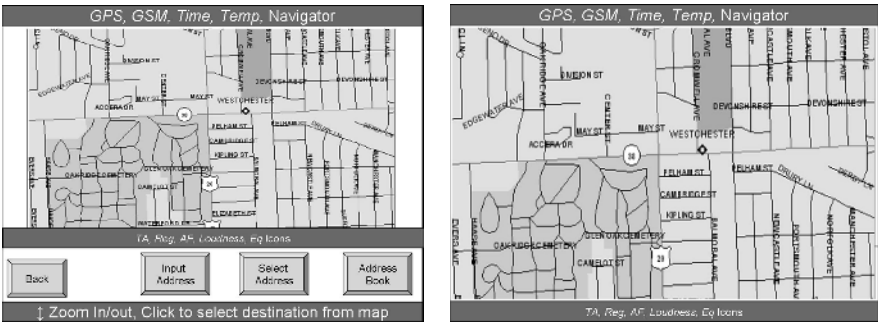
- Notches, to inform about the scrolling among the items of a menu or a list;
- Barrier, to notify the end of a list of items;
- Free movement, to facilitate some scroll operations (for example on the navigation map);

Also the buttons provide a tactile feedback. In particular, if in a specific state there is no function associated to a button, it will result locked.

4 The Graphical User Interface of AD-HMI

To fully exploit and support the features of the *Handy* device, we are designing a specific multimodal interface, able to make use of the visual, auditory and tactile channels. In the following we will present the proposed Graphical User Interface (GUI) while the vocal interface, based on the "*earcons*" approach [14], is under development.

It is worth to pointing out that though in the automotive domain the driver's visual channel should not be overloaded, graphical user interfaces offer several benefits with respect to vocal or tactile ones. The most obvious is that GUIs are the best way to present graphical information, such as a navigator map. Moreover visual presentation allows driver for a "self-pacing" of information, as well as it is the most effective way to represent complex data [13]. On the other hand, a wide adoption of automotive vocal interfaces is currently limited both by some strong constraints about the in-car hardware, able to manage only a restricted auditory interaction, and both by some



(a) **(b)**

Fig. 4. The UI of the navigator in the Interaction Modality (a) and View Modality (b)

recognition problems due to the noisy car environment. As a result, also in the most advanced commercial systems (e.g.: *Jaguar S-Type* or *Alfa Romeo GT*), vocal interaction represent a complementary modality, while the GUI is still the main component of an ITS user interface.

When designing GUI for telematics systems, one of the crucial issues is the amount of information to present for each state. The problem is that too much information leads to a confusing UI. Indeed, user has to devote many visual and cognitive resources to identify the needed data among all the shown on the display, with an unacceptable level of distraction. On the other hand, too few information makes a naïve user unable to effectively exploit the system. It is worth to pointing out that when developing ITSs, special attention should be devoted to make it effective for naïve users, because car's buyers do not want to spend time in training themselves with user manuals. This especially holds for fleet cars, employed by rental services, where a client uses a vehicle only for a very limited amount of time.

To address the issue concerning with the amount of information, the key idea of *AD-HMI* is to provide two different GUI layouts, and to switch between them on the basis of the action carried out by the user:

- If (s)he seems to be about to interact with the system, the GUI displays all the information needed to effectively support the user in his/her task. In particular, to achieve this goal, we adopted the *self-revealing* approach [9], i.e. the interface explicitly reveals to user what functions or items are available, as well as how to act in order to utilize them. As a result, even an unskilled user can easily and effectively exploit *AD-HMI*, relying only on the information provided by the interface. We named this layout *Interaction Modality*. For example, with the navigation module, the system can show the calculated path on the map and all the controls needed to insert/modify the destination (see Fig 4.a)
- If (s)he seems to be not interested in interacting, then the system goes into the *View Modality*, showing only some module-specific information, and hiding all the controls needed to interact. In this way the amount of information displayed by the GUI is reduced, while in the meantime the data are shown with larger fonts. As for the navigation example, in the View Modality, the system will present a full screen view of the calculated path on the map (see Fig 4.b).

Now, the main problem is how to discriminate about the user intentions, for switching between the two modalities. This distinction can be achieved by using another feature of *Handy*, i.e. a proximity sensor. Indeed, in the shell of the device it is possible to install a proximity or contact sensor, using a technology similar to the one adopted for touchpads. In this way it is possible to be aware about the position of the driver's hand with respect to *Handy*: if user places his/her hand on the device, then suddenly the system goes into the *Interaction Modality*. If user retracts the hand, then after a certain amount of seconds, the systems come back into the *View Modality*.

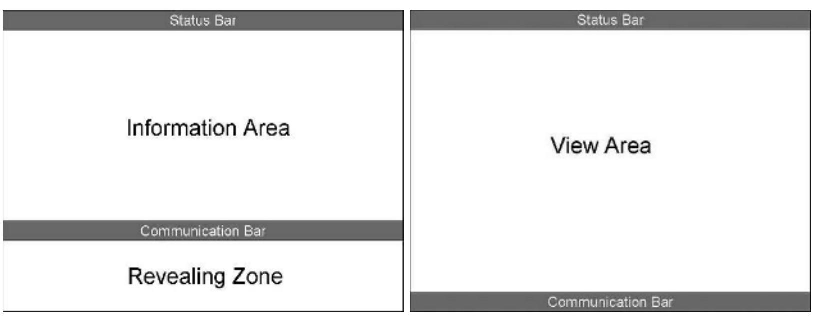
We are starting an analysis to estimate the better timeout value, but preliminary experiments indicates that 20 seconds should be an adequate interval.

4.1 Interaction Modality

The main purpose of *Interaction Modality* is to guide user in the interaction with the system. In particular, to obtain self-revelation, the GUI has to indicate the semantic associated to the *Handy* soft-buttons. A specific section of the GUI, named *Self-Revealing Zone*, is suited to this aim.

Thus, in Fig. 5(a) it is shown the layout characterizing the *Interaction Modality*. As we can see, the GUI is divided in four zones:

- An upper bar, named *Status Bar*, that shows status information about the system, such as the signal level of GSM and GPS, the current time and temperature, and the name of the active module/application (i.e. Navigator, Tuner, Trip Computer, etc…);
- A middle section, named *Information Area*, used to show information related to the active application. For example, with the Navigator module it will show the map, with the Tuner it will show the list of stored radio station, etc…
- A lower bar, named *Communication Bar*, showing the most relevant information of the Phone and Entertainment modules, such as a received SMS, the RDS data, the CD-Text, etc…
- A lower section, the *Self-Revealing Zone*, used to describe the semantic of the actions associated to the Handy buttons, as well as the ones corresponding to the click, tilt and rotation of the scroll wheel.



(a) **(b)**

Fig. 5. Structure of the UI in the Interaction Modality (a) and View Modality (b)

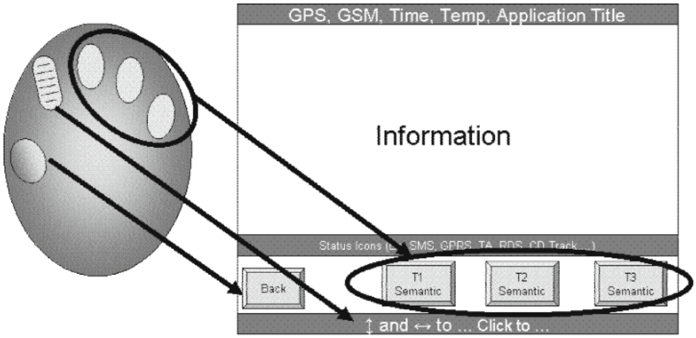


Fig. 6. Mapping between the Handy buttons and the User Interface widgets

In Fig. 6 it is shown how the Self-Revealing Zone informs user about the semantic of Handy buttons and wheel. In particular, such display area consists of four buttons and a bar. The first button, used to describe the TB functionality, is smaller and slightly apart from the others used for T1, T2 and T3 semantic. Finally, a text in the bar describes the functions associated to the movement of the scroll wheel.

Notice that at this stage of the development, we have not defined particular shapes for the controls, neither specific colours, because they will be selected following both the directives in force, and the branding issues.

4.2 View Modality

The View Modality is the standard layout adopted by the GUI during the drive. The aim of this layout is thus to show only relevant information about the current active module. As a result, if the system detects that there is not any imminent user interaction, the *Self-Revealing Zone* is hidden. In the meantime the *Interaction Area* is maximized to magnify the presentation of information (and is now named *View Area*), while the *Communication Bar* is moved on the bottom of the screen. The resulting layout of the GUI is shown in Fig 5(b).

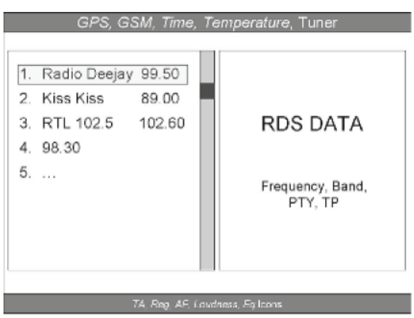


Fig. 7. The UI of the Tuner section in the View modality

The View Area is used to show information about the current active module. For example, with the navigation module, it will contain a full-screen view of the map, centred on the current position of the car, while with the audio module, it should adopt a split screen layout, with the list of the CD tracks on the left and the details of the currently played song, like time, CD Text, etc…, on the right (see Fig. 7).

Finally, in presence of specific asynchronous events, such as an incoming Phone call or SMS, the system pop-ups a dialog box, whose management requires switching to the *Interaction Modality*.

5 Conclusions and Future Work

Safety on the roads is one of the main goals for everyone involved in the automotive field. The advent of ITSs based on visual interaction can distract user from the main task of driving the car, with potentially fatal effects. Nevertheless, visual interfaces offer also evident benefits. Thus, currently special efforts are being devoted to define novel approaches meant to synergistically integrate visual, auditory and tactile information, to reduce the visual workload determined on drivers by the telematics systems.

In this paper we presented the results of a project jointly developed by the research centre Elasis and the University of Salerno, aimed at defining an innovative interface for automotive telematics systems that takes especially into account the safety issues. The main component of the proposal is a new interaction device, named *Handy* and currently patent pending, suited to exploit the driver's tactile channel. Thanks to its specific shape and positioning, the user has always within reach all the commands needed to interact with the system, and can rely on his/her tactile channel to identify the suited controls. Hence, the adoption of *Handy* allows us to significantly reduce the highly distracting task of taking out the glance from the road and looking at the faceplate of the system, with a significant improvement of the overall safety.

In the paper we also presented a graphical user interface able to fully exploit the characteristics of *Handy* and meant to minimize the driver's visual workload. Indeed, to limit the amount of shown information, two different layouts characterize the GUI, depending on the modality of interaction with the system. Moreover, no more than 7 different actions are proposed to the user in any time, thus fitting the typical short-term memory capacity. Another of the distinguishing characteristics of the proposed GUI is the fact that it implements the self-revealing approach, thus making the system suited for naïve users, which was one of the main requirements.

Currently we are realizing a prototype of *Handy* as well as we are implementing the described graphical user interface. This will allow us to conduct extensive evaluation of this proposal on a significant sample of end-users in order to assess the effectiveness of the proposal. We are aware of how much challenging and critical is this issue, because evaluating "in-car" user interfaces requires understanding what effects ITSs may have on driver behaviour and performance. Moreover, the design of *Handy* will require addressing some further issues such as the possibility to invert the buttons layout when the device is placed on the left in a right-hand-drive car together with the comprehension of the consequences on user performance.

References

1. Marcus A., Vehicle User Interfaces: the next revolution, *Interactions*, 1, 2004.
2. Gellatly A., *The use of speech recognition technology in automotive applications,* PhD thesis, Virginia Polytechnic Institute and State University, 1997.
3. Redelemeir D.A., Associations between Cellular-Telephone Calls and Motor Vehicle Collisions. *The New England Journal of Medicine*, Vol. 336, Feb. 1997.
4. Cellario M., Human-Centered Intelligent Vehicles: Toward Multimodal Interface Integration. *IEEE Intelligent Systems*. 2001.
5. Shneiderman B., *Designing the User Interfaces*, Addison Wesley, 1998.
6. *Commission recommendations of 21 December 1999 on safe and efficient in-vehicle information and communication systems: A European statement of principles on HMI*, 1999.
7. Burnett G., Porter J., Ubiquitous computing within cars: designing controls for non-visual use, *Int. J. Human-Computer Studies* (2001) 55, 521-531.
8. Smith E., Driven to Distraction, *New York Times Magazine*, New York, Dec. 2, 2002
9. Kurtenbach G., *The design and the evaluation of marking menus*, PhD thesis, University of Toronto, 1993.
10. Miller G., The magical number seven, plus or minus two: Some limits on our capacity for processing information, Psychological Science, 63, 1956, 81-97.
11. ISO/DIS 2575 Road Vehicles – Symbols for controls, indicators and tell- tales 12/01/95.
12. Rogers, S. P., & Campbell, J. L. (1991). *Guidelines for automobile hand control locations and actuations based upon driver expectancies and ergonomic principles* (TR 947-1). Santa Barbara, CA: Anacapa Sciences, Inc.
13. Ross T., Vaughan G., Burnett G., May A., *HARDIE, Design guidelines for route guidance and navigation*, 1996.
14. Brewster S, Wright P. and Edwards A., An evaluation of earcons for use in auditory human-computer interfaces. *Proceedings of InterCHI'93*, Amsterdam, 1993.
15. Brewster S., Brown L., Tactons: Structured Tactile Messages for Non-Visual Information Display, in *Proc. of 5th Australasian User Interface Conference*, Dunedin, 2004.
16. Gish K., Staplin L., *Human Factors Aspects of Using Head Up Displays in Automobiles: A Review of the Literature*, NTSHC interim report, 1995.
17. *Strategies for Reducing Driver Distraction from In-Vehicle Telematics Devices: A Discussion Document*, Standards Research and Development Branch, Road Safety and Motor Vehicle Regulations Directorate Transport Canada, 2003

Interactive Positioning Based on Object Visibility

Christian Kray and Gerd Kortuem

Computing Department, Lancaster University
Lancaster, United Kingdom
{kray,kortuem}@comp.lancs.ac.uk

Abstract. In this paper we describe a new method and user interface for interactive positioning of a mobile device. The key element of this method is a question-answer style dialogue between system and user about the visibility of nearby objects and landmarks; answers given by the user provide clues about the relative position of the user and allow the verification or falsification of hypotheses about the user's absolute location. This new approach combines the respective strengths of a human user (i. e. fast and reliable object recognition) and a mobile system (i. e. fast computation of numerical data). It enables accurate positioning without requiring any other positioning technologies. A particular advantage of this approach is that it lends itself to the implementation on camera-equipped mobile phones, where it can be used to increase the accuracy of cell-based localisation methods.

1 Motivation

Location-based services (e. g. electronic tour guides [1], location-aware shopping assistance [2]) are one of the key application classes for mobile and handheld devices. Current methods for determining the location (position) of a mobile device (and thus its user) have a number of serious drawbacks in terms of reliability, accuracy, coverage and availability. For example, the quality of location information provided by GPS receivers is effected in unpredictable ways by external factors such as weather and nearby buildings. Essentially all current positioning technologies (including, for example, ultrasound indoor positioning systems, or positioning by detecting mobile phone network cells) have inherent limitations that effect the quality and reliability of the positional information. This fact makes the design of user interfaces for location-based services a challenging task. The basic question is how service fluctuations and disruptions due to technical limitations should be dealt with on the user interface level.

In this paper we propose that rather than trying to hide service fluctuations and disruptions from the user, the user can help to overcome them. We introduce a new interactive positioning method that involves a dialogue between system and user. This dialogue is driven by the system and requires the user to answer a few question regarding the visibility of prominent objects and landmarks (e. g. buildings). The answers provide clues about the relative position of a user with regard to these landmarks and can be used to determine the user's absolute position. The interactive positioning method combines the strengths of a human user – i. e. fast and reliable object recognition – and a mobile system – i. e. fast computation of numerical data – and provides a mean to determine the user's current position even in absence of any sensor readings.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 276–287, 2004.

Of course, asking the user to answer questions can be very disruptive and may lead to an unsatisfactory user experience. Thus crucial research issues related to interactive positioning are *when should the system initiate a dialogue* and *which questions should the system ask?* In this paper, we describe (1) how a dialogue can be generated from a set of initial hypotheses about the user's location (derived from possibly unreliable sensor observations) (2) how knowledge about topology and visibility constraints can be used to select appropriate landmarks and (3) how careful selection of landmarks can be used to minimise the number of questions to be asked.

2 Interactive Positioning

The motivation for employing object visibility for an interactive positioning approach stems from the observation that during mobile phone conversations people often describe their current location, or inquire about the location of another person, in relation to prominent or well-known landmarks (e. g. "I can see a big church with a fountain next to it." "Can you see a statue on a high pillar?"). Our approach takes the idea of dialogues about object visibility and moves it from the realm of computer-mediated human-human interaction to human-computer interaction. It is based on the following key concepts and assumptions:

– Objects (i. e. buildings, landmarks) are visible from a potentially infinite number of positions. Each position is a point in two- or three-dimensional space.
– Given a reference to an object (e. g. a verbal description or a photo), a user can determine whether or not the object is visible from their current position.
– A (finite) number of hypotheses about the user's current location can be derived either using traditional positioning systems or from initial estimates by the user. Each hypothesis refers to a single possible position.

Figure 1 illustrates these concepts. Three objects O1 to O3 are (partially) visible from positions P1 to P3: object O2 is visible from position P1 and P2, whereas O1 is only visible from P2 and neither O1 nor O2 are visible from P3. By asking questions about which objects the user can see (e.g "Can you see objects O1 and O2?") the system can infer whether the user is located at P1, P2 or P3. For example, if the user tells the system that they can see O1, it can infer that the user is in fact located at P2 as this is the only position from which O1 is visible. However, learning that O3 is visible does not rule out any position hypothesis P1 to P3 as it is visible from all of them. It becomes apparent from this example that *for any given situation there is a very large number*

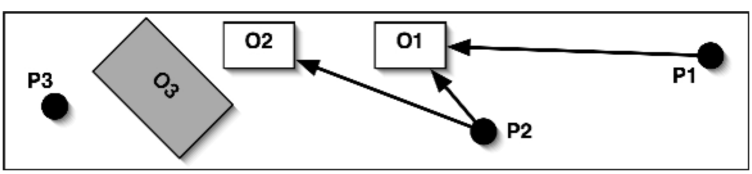


Fig. 1. Varying visibility of objects from different positions: object O1 is visible from position P1 and P2, whereas O2 is only visible from P2 and neither O1 nor O2 are visible from P3

of questions regarding object visibility that the system could ask the user, and *that it is crucial for the effectiveness and user experience to select* good *object, i.e. objects which allow to eliminate many hypotheses once the system knows whether they are visible from the user's current position.* This ability to select good objects is one of the key features of the algorithm we will present in the following section.

2.1 Integration with Existing Positioning Techniques

Most positioning technologies for mobile devices rely on sensors and direct measurements. Figure 2 provides a schematic overview of how interactive positioning can be combined with these. If sensor data is reliable and up-to-date, it provides a precise measurement of the user's current position and no further processing is required. If, however, sensor data is unreliable or outdated, the measurement has a low confidence value. In that case, the system can improve confidence by asking the user for explicit confirmation. For example, it can use a personalised you-are-here map containing well-known landmarks [3] to enable the user to verify the position. In case of no sensor data, we can resort to exploration [4] (i. e. by querying other knowledge sources or by asking the user for some rough estimation of their position such as "What quarter of the city are you in?"). While confirmation and exploration are a form of interactive positioning, our approach goes beyond simple dialogues by means of an optimised interaction with the user regarding the visibility of objects. The combination of direct measurement, inference and interactive positioning allows for the determination of the user's current position in a wide array of situations, where individual techniques would fail. In particular, our approach is even able to estimate the user's position without any sensor data at all (e. g. if the user provides some rough initial estimate such as "I am on a market square.").

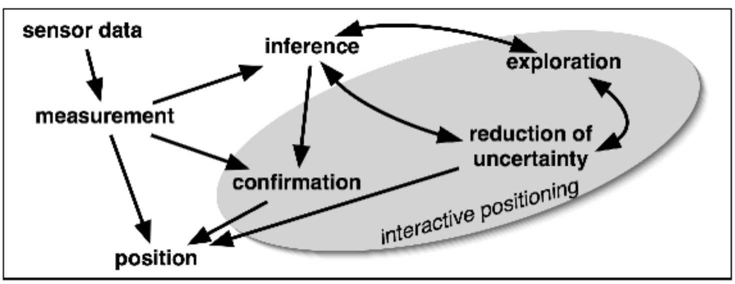


Fig. 2. Determination of the user's current position: an overview

2.2 Technical Requirements

In order to realise interactive positioning, a mobile system has to meet two key requirements. First, it needs to provide a user interface suitable for interactive dialogues between system and user. This may be done visually using a graphical user interface (GUI) or verbally using speech synthesis and speech recognition. Second, it needs to have access

to a geographic information system (GIS). This GIS must contain a world model that allows for the computation of object visibility from arbitrary positions.

In the following section we describe how to generate a suitable dialogue from basic hypotheses. The discussion assumes a graphical user interface with the capability of displaying photos of individual objects and slideshows of sets of objects. A concrete example is given in Section 4.

3 Generating Dialogues for Interactive Positioning

Interactive positioning based on object visibility is an iterative process that takes as input a number of hypotheses (e. g. generated from imprecise measurements or through dead reckoning), and calculates a number of questions concerning the visibility of objects in the surroundings. These questions are optimised towards quickly determining the current position of the user. Figure 3 shows an overview of the entire algorithm that reduces the uncertainty, which position of a set of several candidates is most likely the actual position of the user. We first select the best divider, i. e. the salient object that partitions the visibility matrix (see below) in a way that allows us to quickly reduce the size of the matrix once we know whether the object is visible. In order to limit the number of interactions, we select the best dividers for the resulting sub-matrices as well. Then, we generate the query for the user, which consists of a repeating slide show of labelled images of the selected salient objects. The user's reply reduces the matrix according to the procedure described below. The algorithm terminates either when the user's position has been determined, or when it cannot identify it. In the latter case, we can resort to exploration (see Section 2).

```
select best dividers
    find salient object that best divides matrix
    find salient objects that best divide submatrices
generate query for user
    retrieve images for salient objects
    show repeating slide show and question
evaluate the user's reply
    reduce matrix
    if matrix is empty
       exploration
    else if matrix has only one element
       return as user's position
    else if elements of matrix can be merged
       return as user's position
    else if matrix does not allow for further reduction
       exploration
    else repeat
```

Fig. 3. The reduction algorithm: an overview

$$V(S,P) = \begin{pmatrix} vis(s_1,p_1) & \dots & vis(s_1,p_n) \\ vis(s_2,p_1) & \dots & vis(s_2,p_n) \\ \vdots & \ddots & \vdots \\ vis(s_m,p_1) & \dots & vis(s_m,p_n) \end{pmatrix}$$

$$S = \{s_i | 0 < i < m+1\}$$

$$P = \{p_j | 0 < j < n+1\}$$

$$vis(s_i,p_j) = \begin{cases} 1 \text{ iff } s_i \text{ is visible from } p_j \\ 0 \text{ otherwise} \end{cases}$$

Fig. 4. Visibility matrix and its constituents: the set S of salient objects s_i, the set P of all potential positions p_j, and the visibility function $vis(s_i, p_j$.

In a first step, the algorithm retrieves all world objects that are close to the potential positions. In order to reduce the number of objects, a preselection based on their respective salience (see, for example, [5]) should be performed. For all the *salient objects* of the resulting set S, we then determine whether or not they are visible, i. e. for all potential positions we check the visibility of each object. We then dispose of a *visibility matrix* $V(S,P)$ as shown in Figure 4, the central data structure for the algorithm.

Since the user's reply to a question – whether or not they can see an object – should allow us to eliminate as many hypotheses as possible, we have to select those salient objects that best partition the set of the potential positions. An ideal example for such an item would be a salient object that is visible from exactly half of the potential positions. Usually, there is no such object, and we have to instead select the ones that partition the set of potential positions in two sets of roughly the same size. More formally, we are looking for the salient object s_k (see Figure 4 for the definition of the terms used) for which the following statement holds:

$$s_k : \forall i \in \{1,\dots,k-1,k+1,\dots,m\} :$$
$$\left|\tfrac{n}{2} - \textstyle\sum_{j=1}^{n} vis(s_i,p_j)\right| \leq \left|\tfrac{n}{2} - \textstyle\sum_{j=1}^{n} vis(s_k,p_j)\right|$$

If more than one salient object meets this criterion we can either randomly select one, or recursively determine which salient object entails the lowest number of questions once its visibility is known. The latter alternative yields a more informed choice (at the expense of higher computational costs), since we analyse *in advance* what questions will follow when the user either confirms visual contact with the current salient object or not. This approach can also be iterated after each reply by re-evaluating the set of the remaining candidates in the same way (again at the expense of higher computational costs). Note that this approach ensures that the number of interactions required to determine the user's current position is minimised as it always selects those objects that result in the highest expected information gain.

$$elim_s(k,V) = \begin{pmatrix} vis(s_1,p_1) & vis(s_1,p_2) & \dots & vis(s_1,p_n) \\ \vdots & \vdots & \ddots & \vdots \\ vis(s_{k-1},p_1) & vis(s_{k-1},p_2) & \dots & vis(s_{k-1},p_n) \\ vis(s_{k+1},p_1) & vis(s_{k+1},p_2) & \dots & vis(s_{k+1},p_n) \\ \vdots & \vdots & \ddots & \vdots \\ vis(s_m,p_1) & vis(s_m,p_2) & \dots & vis(s_m,p_n) \end{pmatrix}$$

$$elim_p(j,V) = \begin{pmatrix} vis(s_1,p_1) & \dots & vis(s_1,p_{j-1}) & vis(s_1,p_{j+1}) & \dots & vis(s_1,p_n) \\ vis(s_2,p_1) & \dots & vis(s_2,p_{j-1}) & vis(s_2,p_{j+1}) & \dots & vis(s_2,p_n) \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ vis(s_m,p_1) & \dots & vis(s_m,p_{j-1}) & vis(s_m,p_{j+1}) & \dots & vis(s_m,p_n) \end{pmatrix}$$

$$V' = elim_s(x, V_k) \text{ where } k = \sum_{j=0}^{n} vis(s_x, p_j)$$

$$V_l = \begin{cases} l = 0 : V \\ l \neq 0 : elim_p(min(j | j < columns(V_{l-1}) \wedge vis(s_x,p_j) = 0), V_{l-1}) \end{cases}$$

Fig. 5. Elimination of false hypotheses in the visibility matrix

Once the user provides the system with visibility information (either for one salient object or for several), we can adjust the visibility matrix by eliminating all positions that contradict the user's reply. The row(s) corresponding to the salient object(s) included in the query can be removed as well since it is of no further use. Figure 5 shows the formal procedure of elimination for a single salient object question. (Multiple salient objects questions can be treated as a sequence of single salient object questions.)

In order to determine the updated visibility matrix V' we need to eliminate all potential positions from the original matrix V from which salient object s_x is not visible. This is an iterative process that removes position p_j if $vis(s_x, p_j) = 0$ resulting in intermediate matrices V_l. Once all invisible positions have been eliminated, the current salient object s_x can be removed as well ($elim_s(x, V_k)$). If the user reports that s_x is invisible, the only differences are that we have to eliminate positions p_j if $vis(s_x, p_j) = 1$, and that $k = n - \sum_{j=0}^{n} vis(s_x, p_j)$. If the user provides information about multiple salient objects simultaneously we can apply the same procedure to one salient object after the other. The process of interaction and elimination continues until

- **all hypotheses have been eliminated.**
  This implies that either the original set of potential positions was wrong, or that the user was unable to recognise a salient object, or has overlooked one or more salient objects.
- **there is only one hypothesis left in the visibility matrix.**
  We have successfully determined the user's current position, and the system can proceed with the task that requested positional information.
- **the remaining hypotheses can be merged into a single position.**
  This happens when the remaining salient objects do not allow for a reduction of

uncertainty (e. g. they are visible from all positions), and if the remaining hypotheses are located close to each other.

– **the remaining hypotheses cannot be merged.**
  In this case, the remaining salient objects cannot be used to reduce the uncertainty, which of the remaining positions is the the true position of the user, and they are also too far apart to be merged.

The second and third case allow for the termination of the positioning process, and enable the system to continue to work on the task that originally requested information about the user's current position. In the first and fourth case, however, the reduction of uncertainty has failed, and other means have to be employed to still provide the service the user has asked for. A convenient possible 'by-product' of the interaction on the visibility of objects is a hypothesis about the current orientation of the user: in order to answer visibility questions, they mostly likely will look towards the objects in question and align themselves accordingly. A further beneficial side effect of the interaction is the introduction of a number of world objects that the system can later refer to, e. g. when generating localisations.

4 Case Study

We now describe a case study of using interactive positioning in a mobile tourist guide. We first illustrate the user experience when interacting with the prototype system and present some qualitative and quantitative results both from lab tests and a field trial.

4.1 User Interface

Figure 6 shows a snapshot of the user interface of Deep Map, a mobile tourist guide that uses interactive positioning. Deep Map provides visitors of the city of Heidelberg with a number of location-based services. The upper half of the figure depicts an example interaction: The system asks the user, which of three objects are visible from their current position. In order to facilitate recognition, a slideshow is presented to the user that consists of photographs of the objects that are annotated with their name. The slideshow is repeated until the user replies to the query. This does not only enable the user to identify the objects in their environment but also provides an easy mean to refer to them when replying to the system. The lower half of the picture illustrates the actual context of this interaction: the user is currently located on Seminarstraße and the objects mentioned in the query are nearby that street (all are highlighted in the map).

The interface shown in Figure 6 is one step in a (short) series of questions that occur during interactive positioning. Once the user provides the system with a reply – in our implementation via a pop-up menu and/or an on screen keyboard – it can update the visibility matrix and generate the next question if necessary. Once Deep Map has successfully determined the current position, it will update the internal position history and provide the user with its location-based services, e. g. personalised you-are-here maps such as shown in Figure 7.

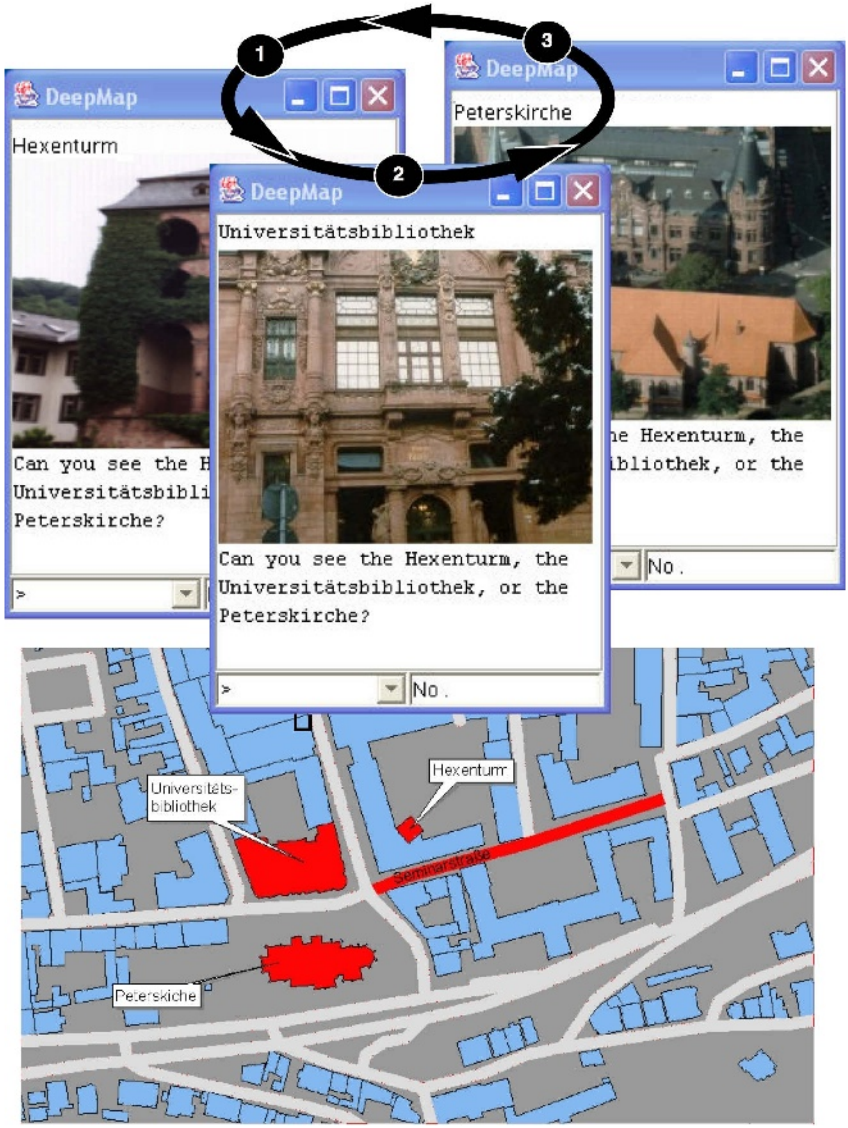


Fig. 6. An example interaction: The user is located somewhere on Seminarstraße (highlighted on the map), and the system now asks whether 'Hexenturm' (1), 'Universitätsbibliothek (2), or 'Peterskirche' (3) are visible (also highlighted on the map). Images of these three objects are shown in a continuously repeating slideshow (indicated by the circular arrows at the top of the figure) along with their name. In the prototype, the user could reply to the question in two ways: either by using the pop-up menu in the lower left hand corner of the screen/window to select a predefined answer (such as "Yes." or "No.") or by inputting free text into the input box in the lower right hand corner.

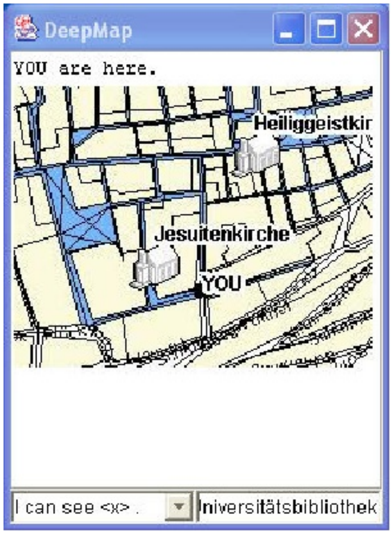


Fig. 7. A personalised you-are-here map that was generated after successfully applying interactive positioning (the objects shown on the map are known to the user).

4.2 Evaluation

Our prototype is an extension of the Deep Map system [6], a system that provides services such as incremental navigation, information on sights, hotel reservation, and interactive maps. Deep Map relies on GPS to determine the user's current position but has been designed to easily integrate other positioning techniques.

Deep Map was tested during development within the lab using a GPS simulator agent that allowed us to simulate accurate measurements as well as the complete lack of any readings. In those tests, we found the system to be able to determine the current position in a number of different situations, ranging from the complete absence of positional information to a set of different position hypotheses. However, on open places (such as market squares or wide roads), we observed a higher number of cases, where the system was not able to pinpoint the user's position beyond a relatively low precision. We attribute this to the implementation of the visibility check: The algorithm used to compute visibility was based on a two-dimensional ray-tracing approach and therefore could not evaluate the visibility of objects that are further away but tall.

In addition to lab tests we also conducted a field trial with the system. In this case, we disabled the GPS while the user was on Seminarstraße (see also Figure 6) so that the system did not have any current sensor data. This street is approximately 170 meters long, and the prototypical implementation was able to determine the user's current position in three interactions such as the ones shown in Figure 6. The computed position was accurate within an ten meter radius.

5 Related Work

There are a few systems – mainly prototypical tourist guides or navigational assistants – that already incorporate interactive positioning in one form or another. The GUIDE project [1] was developed at the University of Lancaster, and aims at providing visitors of the city with information adapted to their interest and location. GUIDE can present its user with a list of all sights from which they can pick the one that is located nearest to them. Based on this selection, the system can then estimate the user's current position. While the list of sights GUIDE presents to the user is static, the LoL@ system [7] is able to dynamically generate it. LoL@ is also aimed at tourists, who it provides with information about points of interest, navigational assistance, and further location based services. Currently, it relies on GPS for positioning but it has been designed to exploit the position information provided by third generation mobile phones. In case LoL@ is unable to precisely determine the user's current position from sensor readings or through dead reckoning, it dynamically creates a list of street segments and ask the user to select the one they are located on. This list consists of ranges of house numbers along with the name of the street. Hence, this approach requires LoL@ to know the street the user is on.

A further interaction technique used for positioning consists of interactive maps, where the user can 'point' to their current location by clicking on the corresponding area on the screen of a PDA. Within the project REAL [8], for example, the imprecision of positional information is compensated by displaying a larger area of the environment. The user can then click on specific icons embedded in the map to tell the system about their current location. This information is then used to improve the quality of the presentation, i. e. by providing more precise route instructions. Bhasker et al. [9] use a similar approach to improve the precision of WLAN-based positioning but store user corrections as 'virtual access points' for later use.

However, there are several shortcomings in the approaches presented above. A static list of sights does not scale well – in a larger city, a user might have to select from thousands of items – and also restricts the precision of the resulting positional information. A dynamically generated list of street segments overcomes this problem to some degree but does require information about the street the user is in. In addition, longer streets will result in a long list of street segments, which are in turn hard to communicate to the user on a mobile device with limited screen estate. Interactive maps enable the user to quickly communicate their current location to the system but not only do they have to know their position rather precisely but they must also be able to indicate it on a map.[1]

The approach proposed in this paper addresses these issues in several ways. The iterative nature of the algorithm allows for a fine-grained control of the number of objects to present to the user. In addition, the objects are selected to maximise the expected information gain - hence, once their visibility is known the number of remaining alternatives is drastically reduced. Furthermore, contrary to interactive maps, our algorithm

[1] An alternative approach to 'interactive positioning' consists of adapting the interactions in the context of services to low-precision positional information instead of trying to pinpoint the user's position more precisely (cf. e. g. [10]). However, there is a minimum precision for most services, which has to be met in order to provide them at all.

will work even if the user does not have any idea about their current location. It only expects the user to be able to visually scan their environment and to recognise objects that are presented to them.

6 Discussion

In our case study, we put the 'burden' of checking the visibility of objects on the human user but our approach would also support a system-side check of visibility. This opens up an interesting application area for mobile phones that are equipped with a camera. Instead of going through a number of interactions in order to determine their current position, a user could simply take a few snapshots of their actual environment and send them to a server. The server could then perform image analysis to match the photographs with others that are stored in its geo-referenced database. This does then provide the visibility information needed to reduce the visibility matrix according to the algorithm presented in Section 3. A further advantage of applying our approach to mobile phones with a camera is the fact that the current network cell of the phone provides an initial seed for constructing the visibility matrix. Consequently, the search space for the image analysis is also reduced to those images that are linked to the area of the current cell.

7 Conclusion and Outlook

Designing user interfaces for location-based services is a challenging task due to the inherent limitations of traditional positioning technologies in terms of reliability, accuracy and coverage. In this paper we proposed a new method and user interface for interactive positioning that is able to overcome these limitations. The interface uses a system-driven dialogue to resolve question regarding the visibility of prominent objects and landmarks; answers given by the user provide clues about the relative position of the user and allow for the verification or falsification of hypotheses about the user's absolute location.

In this paper, we described a method for generating a dialogue from basic hypotheses and we demonstrated how interactive positionig based on object visibility can be integrated into a mobile tourist guide system. Unlike previous approaches, our approach dynamically adapts the interaction to maximise the information gain from each interaction step while minimising the length of the interaction. The proposed mechanism not only allows one to specify how precisely the position of the user has to be determined but also seamlessly integrates with non-interactive approaches. A particular advantage of our approach is that it lends itself to an implementation on camera-equipped mobile phones where it can be used to increase the accuracy of cell-based localisation methods

Acknowledgements. The work presented in this paper was funded by the Klaus Tschira Foundation, Heidelberg, Germany in the context of the project SISTO.

References

1. Cheverst, K., Davies, N., Mitchell, K., Friday, A.: Experiences of developing and deploying a context-aware tourist guide: the GUIDE project. In: Mobile Computing and Networking, Boston, U.S.A. (2000) 20–31
2. Bohnenberger, T., Jameson, A., Krüger, A., Butz, A.: Location-Aware Shopping Assistance: Evaluation of a Decision-Theoretic Approach. In Goos, G., Hartmanis, J., van Leeuwen, J., eds.: Proceedings of Mobile Human-Computer Interaction 2002, New York, Heidelberg, Berlin, Springer (2002) 155–169
3. Sorrows, M.E., Hirtle, S.C.: The nature of landmarks in real and electronic spaces. In Freksa, C., Mark, D.M., eds.: Spatial Information Theory (Proceedings of COSIT 99), Berlin, Heidelberg, New York, Springer (1999) 37–50
4. Kray, C.: Situated Interaction on Spatial Topics. PhD thesis, Computer Science Department, University of Saarland, Saarbrücken, Germany (2003)
5. Lynch, K.: The Image of the City. MIT Press, Cambridge, MA (1960)
6. Malaka, R., Zipf, A.: Deep Map - challenging IT research in the framework of a tourist information system. In Fesenmaier, D.R., Klein, S., Buhalis, D., eds.: Information and communication technologies in tourism 2000. Springer, Berlin, Heidelberg, New York (2000) 15–27
7. Anegg, H., Kunczier, H., Michlmayr, E., Pospischil, G., Umlauft, M.: LoL@: designing a location based UMTS application. Elektrotechnik und Informationstechnik **119** (2002) 48–51
8. Butz, A., Baus, J., Krüger, A., Lohse, M.: A hybrid indoor navigation system. In: Proceedings of IUI 2001, New York, ACM Press (2001) 25–32
9. Bhasker, E., Brown, S.W., Griswold, W.G.: Employing user feedback for fast, accurate, low-maintenance geolocationing. Technical Report CS2003-0765, UC San Diego, San Diego, CA, USA (2003)
10. Sood, S., Hammond, K.J., Birnbaum, L.: Low-fidelity location based information systems. In Nunes, N.J., Rich, C., eds.: Proceedings of IUI 04, New Yok, NY, USA (2004)

IDeixis – Searching the Web with Mobile Images for Location-Based Information

Konrad Tollmar[1,2], Tom Yeh[1], and Trevor Darrell[1]

[1] MIT CSAIL, 200 Technology Square, Cambridge, MA 02139
{konrad, tomh, trevor}@csail.mit.edu
[2] Department of Informatics, Lund University, SE-223 63 Lund, Sweden.

Abstract. In this paper we describe an image-based approach to finding location-based information from camera-equipped mobile devices. We introduce a point-by-photograph paradigm, where users can specify a location simply by taking pictures. Our technique uses content-based image retrieval methods to search the web or other databases for matching images and their source pages to find relevant location-based information. In contrast to conventional approaches to location detection, our method can refer to distant locations and does not require any physical infrastructure beyond mobile internet service. We have developed a prototype on a camera phone and conducted user studies to demonstrate the efficacy of our approach.

1 Introduction

Location-based information services offer many promising applications for mobile computing. While there are many technologies for determining the precise location of a device, users are often interested in places that are not at their exact physical location. There are no common or convenient means to make a pointing (deictic) gesture at a distant location with existing mobile computing interface or location-based computing infrastructure.

In this paper, we present Image-based Deixis (IDeixis), an image-based approach to specifying queries for finding location-based information. This is inspired by the growing popularity of camera-phones and leverages the fact that taking a picture is a natural and intuitive gesture for recording a location. The key idea is that with a camera phone, users can point at things by taking images, send images wirelessly to a remote server, and retrieve useful information by matching the images to a multipurpose database such as the World Wide Web. Here we describe a scenario to illustrate an instance of the general idea:

"Mary is visiting campus for the first time ever. She is supposed to meet a friend at "Killian Court". She is uncertain if this place is the "Killian Court". She takes an image of the building in front her and sends it to the server. This image is then used to search the web for pages that also contain images of this building. The server returns the 5 most relevant web pages. By browsing these pages, she identifies the names 'Killian Court' and 'The Great Dome' and concludes that this is the right place."

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 288–299, 2004.

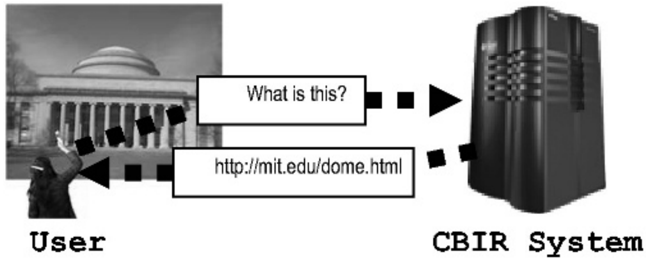


Fig. 1. Mobile Image-based Deixis.

Our system consists of two major components: a client-side application running on a mobile device, responsible for acquiring query images and displaying search results, and a server-side search engine, equipped with a content-based image retrieval (CBIR) module to match images from the mobile device to pages in a generic database (Figure 1).

We first review related work in the literature, and describe the findings of an interview study on user needs for location-based information services. We then describe a prototype constructed to demonstrate the technical feasibility of the concept of the image-based deixis. Finally, we report the results of a user study using a second prototype designed to test and compare different interface approaches.

2 Related Work

Image-based deixis touches four related areas: augmented reality with camera equipped mobile devices, content-based image retrieval, location recognition in robotics and wearable computing, and location-based information retrieval.

Camera-equipped mobile devices are becoming commonplace and have been used for a variety of exploratory applications, e.g. Mobile image matching and retrieval has been used by insurance and trading firms for remote item appraisal and verification with a central database [3]. Other examples are, the German AP-PDA project built an augmented reality system on camera-equipped iPAQ [6], FaceIT ARGUS [5] and a pen-size camera to capture images of text the developed at HP [16]. These systems are successful cases of information retrieval made possible by camera-equipped mobile devices, but they require specific models (e.g. for appliances) and are unable to perform generic matching of new images.

Content-based image retrieval systems can perform generic image matching and have been developed in the past decade for the application of multimedia data mining and archival search. One of the first such systems was IBM's Query-By-Image-Content (QBIC) system [13]. It supported search by example images, user-drawn pictures, or selected color and texture patterns, and was applied mainly to custom, special purpose image databases. In contrast, the Webseek system [19] searched generically on the World Wide Web for images and incorporated both keyword-based and content-based techniques. The Diogenes system used a similar dual-modal approach for searching images of faces on the web [2]. To this day, these systems have not been applied to the task of recognizing locations from mobile imagery.

The notion of recognizing location from mobile imagery has a long history in the robotics community, where navigation based on pre-established visual landmarks is a well-known technique. The task of simultaneously localizing robot position and mapping the environment (SLAM) has received considerable attention [12,10]. Similar tasks have been addressed in the wearable computing community, with the goal of determining the environment a user is walking through while carrying a body-mounted camera [22]. Closely related to our work is the wearable-museum guiding system built by Schiele and colleagues that utilizes a head-mounted camera to record and analyze the visitor's visual environment [17]. In these robotics and wearable computing systems, recognition was only possible in places where the system has physically been before.

There are already location information services offering on-line maps (e.g. www.mapquest.com), traffic reports, and marketing opportunities on mobile devices. An often-discussed commercial application of location-based information is proximity-based coupon delivery. In a typical scenario, a merchant is notified when a potential customer visits a nearby retail outlet, upon which the customer can be delivered a coupon or offered a special promotional deal. The comMotion project [11] extended this idea to allow user-side subscription-based and location-specific content delivery. The GUIDE system was designed to provide city visitors location specific information customized to their personal needs [4]. The identity of the location is provided by the underlying cell-based communication infrastructure, which is also responsible for broadcasting relevant tourist information.

In [9], Kaasinen examined the usage of location-based information through interviews. This study highlighted the need for more comprehensive services in terms of geographic coverage, variety (number of services offered) and depth (amount of information available).

3 IDeixis - Image-Based Deixis

IDeixis is intended to be a pointing interface paradigm and location-based computing technique which combines the ubiquity of a new generation of camera-phones with CBIR and the world wide web. There are two main components to a location-based computing system: the specification or sensing of location, and querying a database for location-relevant content. We will discuss the image-based approach to each in turn.

In our IDeixis system users specify a particular location by pointing to it with a camera and taking images. The location can be very close, or it can be distant – it just must be visible. IDeixis allows users to stay where they are and point at a remote place in sight simply by taking photographs. IDeixis does not require any dedicated hardware infrastructure, such as visual or radio-frequency barcode tags, infrared beacons, or other transponders. No separate networking infrastructure is necessary besides what is already made available by existing wireless service carriers, for example, General Packet Radio Service (GPRS) and Multimedia Messaging Service (MMS).

Having specified a location, a location-based information service needs to search for geographically relevant messages or database records. While geographic cues may well become a standard metadata tag on the web, they are not at present commonly

available. However one form of location cue is already ubiquitous throughout the internet—images of actual places. If we can develop a method to match images from mobile cameras to images on the internet, we can gather web pages based on the geographic location of the camera-equipped device.

We believe the wealth of information already contained in the web can be exploited for location-based information services. Indeed, keyword-based search engines (e.g. Google) have established themselves as the standard tool for this purpose when working in known environments. In practice, current web-image search engines, such as Google, use keywords to find relevant images by analyzing neighboring textual information such as caption, URL and title [1][7]. However, formulating the right set of keywords can be frustrating in certain situations [8]. For instance, when the user visits a never-been-before place or is presented with a never-seen-before object, the obvious keyword, name, is unknown and cannot be used as the query. In our work, this process is reversed. An image is used to find matching images of the same location. In many situations, finding these images on the web can lead us to the discovery of useful information for a particular place in textual form.

Image-based deixis could be desirable in this situation: the intent to inquire upon something is often inspired by one's very encounter of it - the place or object in question is conveniently situated right there. With a camera phone, an image-based query can be formed simply by pointing with the camera and snapping a photo.

4 First Prototype

A pilot prototype was designed to find out whether searching for location-based information on the World Wide Web by matching images from a mobile device is practical and useful. We tested whether current CBIR algorithms can be effectively applied to images of locations and how feasible it was to implement a real system with existing camera-equipped mobile devices and wireless network infrastructure.

4.1 System Design

We built our first prototype on a Nokia 3650 phone taking advantage of its built-in camera (640×480 resolution) and the support for Multimedia Messaging Service (MMS), using C++ on Symbian OS [21]. To initiate a query, the user points the camera at the target location and takes an image of that location, which is sent to a server via MMS.

We designed our system with an interactive browsing framework, to match users' expectations based on existing web search systems. For each query image, the search result will contain the 16 most relevant candidate images for the location indicated by the query image. Selecting a candidate image brings up the associated web page. The user can browse this page to see if there is any useful information.

4.2 Experiment and Results

The first prototype was built to evaluate whether CBIR can match location images from mobile devices to the pages on the world wide web. For our initial experiments,

we restricted ourselves to a known domain, a single university campus, both for web searching and when initiating mobile queries. We began by constructing an image database consisting of 12,000 web images collected from the mit.edu domain by a web crawler. Test query images were obtained by asking student volunteers to take a total of 50 images from each of three selected locations: Great Dome, Green Building and Simmons Hall. Images were collected on different days and with somewhat different weather conditions (sunny/cloudy); users were not instructed to use any particular viewpoint when capturing the images.

We tested the image matching performance of two simple CBIR algorithms: windowed color histogram [20] and windowed Fourier transform [15]. Principal component analysis was used for finding the closest image in terms of Euclidean distance in the feature space. These are among the simplest CBIR methods in the literature, and partial success with these methods would suggest even greater performance with more advanced techniques (this is discussed in more depth in [23]).

These volunteers were used as test subjects to evaluate the performance of these algorithms. For each test query, the search result consisted of the 16 closest images in the database to the query image using the selected distance measure (color histogram or Fourier transform). We asked the test subjects to decide how many of these candidate images are actually similar to the original query images. Figure 2 summarizes the performance we obtained from the tested image matching metrics. It shows the percentage of tries our test subjects found at least one similar image on among the first 16 candidate images.

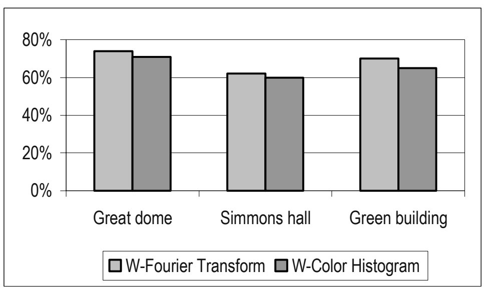


Fig. 2. Average retrieval success for each CBIR algorithm shown in percentages of 150 test attempts that returned at least some relevant images on the first page of candidate images.

Unfortunately, the underlying MMS-based communication infrastructure had in practice an average turnaround time of 25sec, which required much patience of the user. We expect as MMS becomes more popular, with better technology implemented by wireless service carriers, the turnaround time can be drastically improved. We are also currently exploring image compression methods and better network protocols to improve interactive performance.

4.3 Discussion

Analysis of the example webpages found in the study (Figure 3) shows that we can in principle find web pages with images that perfectly match the image of the current

location, yet the web page that contains that image may be poorly related to the location – for example, a photo gallery of someone who visited a museum rather than the official page of that museum. With an interactive paradigm, users can keep searching for a more appropriate match but this may take considerable time. Additionally, we observed that people often searched for more specific questions beyond "what is that". They needed to go back and forth between the thumbnail mosaic and the web browser to examine many web pages for a specific piece of information. To do so on the small-screen device can be too cumbersome to be feasible.

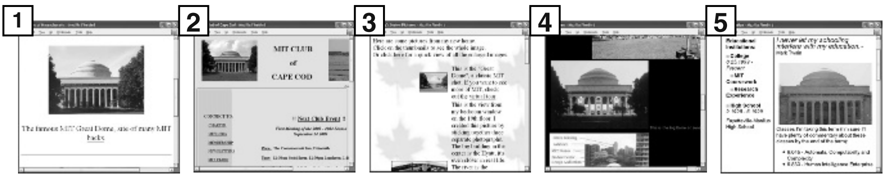


Fig. 3. Some examples of found webpages: (1) MIT Gallery of Hacks, (2) MIT Club of Cape Cod's official website, (3) Someone's picture gallery, (4) Someone's guide on Boston, and (4) Someone's personal homepage.

5 Interview Study

In order to better understand the browsing after location-based information using mobile devices we conducted a interviews study before we designed the second prototype. The interview study involved 20 subjects and took place in common tourist locations. First we went through a list of questions about location-based information. Then, we accompanied subjects as they walked around and encouraged them to talk out loud about what they were seeing and describe the kind of information the would appreciate in that context. We also handed the subject a camera and asked them to take pictures of the objects of interest. The data collected allowed us to address the following two questions:

- How do people currently use maps and tour books while visiting an unfamiliar location?
- What do people want to know about their specific location, and how do they want to obtain the information?

Regarding the first question, we were told that maps and tour books often lack detailed information. Very few subjects reported bringing them in everyday life. But virtually everyone reported carrying a map when traveling to a new place. One interesting finding was the tendency of people to overstate the usefulness of a street map. On second thought, they would retract such statement as they quickly realized they actually wanted to know more than what a map could provide, such as specific details about buildings and artifacts they were seeing around them.

Based on our observations, we found that there are many specific questions asked only by certain individuals, like "what kind of bike is this", "what is the name of this

tree", and "when does the city empty these garbage bins." However, there were also general questions shared by most of the subjects. These questions range from historic information and events to names of buildings and makers of public artworks. We identified the two most commonly asked questions were "Where can I find X?" and "What is this?" Often times, these questions were followed by requests for time-related information such as business hours and bus schedules.

From these interviews, we concluded that location-based information services which provided access to a generic information service such as the World Wide Web, and which were initiated by a real-time query (e.g., "What is this place?") followed by a browsing step, would most often complement the users' experience in an unfamiliar setting and meet their needs for a location-based information service. But we also observed that that people often searched for more specific questions beyond "What is this?".

6 Second Protype

To match these more specific searches for information, we developed a new "keyword bootstrapping" approach in which image-based keywords are used for interactive searching. The first steps of the process are as before, with a user taking a picture of a location, and image search returning a set of matching images and associated web pages. Salient keywords are automatically extracted from the image-matched web pages. These keywords can then be submitted to a traditional keyword-based web search. In the museum example, the museum name (e.g. "Louvre") would presumably appear as a keyword on many of the matched pages, and a Google keyword search with "Louvre" would very likely yield the museum's official homepage. With this approach the relevant homepage can be found even when it contains no image of the location itself.

Our second prototype explored the image-based keyword bootstrapping paradigm and compared it to the thumbnail browsing approach of the first prototype. We also compared performance to a baseline map-based method where users can click on a desired location on a small map to find information. We used the same hardware as the first prototype but used a web interface developed in XHTML Mobile Profile with JavaScript extension [14]. We implemented three search strategies:

- **Searching for web pages:** the search result consists of a list of matched web pages containing similar images of the query image. Each page is displayed as a thumbnail accompanied by a text abstract of its content. Selecting a thumbnail brings up the full content of the page on the screen (see Figure 5).
- **Searching for keywords:** Automatically extracted keywords are displayed side-by-side with the thumbnail image. Selecting a keyword initiates a keyword-based search on Google to find more information (see Figure 5).
- **Searching by mobile MapQuest:** We used a GPS-coordinate-based query to retrieve from MapQuest a map covering the surrounding area. This is meant to serve as the baseline for evaluating image-based approaches.

To compare the usability of these three approaches without the issue of network latency, we pre-computed the image search and pre-cached portions of the web interface and result pages for known landmarks. In this part of the study, we focused on the relative ease of use of the three interfaces rather than evaluating the matching success of the CBIR algorithms.

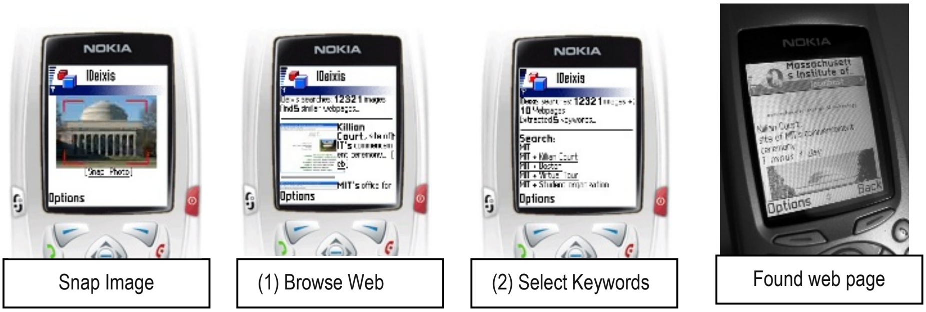


Fig. 4. Two image-based search strategies for location-based information. First strategy (1) displays the results directly on the screen in thumbnails. Second strategy (2) provides a set of keywords for a user to select and submit to a text-based search engine.

6.1 Method

We conducted the testing of our prototype at two different outdoor sites under fairly good weather conditions with 16 subjects aged between 13 and 63 evenly split between genders. A survey on their background of the technology revealed that all of them use the Internet regularly. Most of them (14 out of 16) owned a cell-phone; 11 used it as their primary phone. Some (6 out of 16) owned a digital camera; 4 used it frequently. The testing of the prototype was conducted in two steps:

1. We handed the subject a Nokia 3650 camera phone and asked her to walk around taking pictures of any particular landmark on the site she would like to know more about.
2. We let the subject use and evaluate the prototype system on the phone to perform the task of searching location-based information, using each of the three search strategies.

All the testing sessions were recorded on video. We later analyzed the video sequence manually and extracted from it the time it took to locate a piece of information at each search attempt as well as the subjective evaluation of the quality of information found by the subjects (see Table 1).

6.2 Results

All our subjects found it very intuitive to express their interests in a particular place by taking pictures in an attempt to look for location-based information. This gives us

substantial, if not conclusive, evidence to claim that IDeixis is indeed a simple and intuitive way to specify the locations of interest.

Although all of our subjects were familiar with the Web, none of them had any experience in surfing the Web on a mobile device. Nonetheless, basic web browsing on the device was not a problem. On the contrary, we got several very possitive feedbacks, such as: "*Internet in the phone is cool*".

On searching for web pages: We observed among subjects an interesting disparity in their perception of how specific search results were. One subject commented upon this by reflecting that "… *similar web pages are very different in content*"; he thought the search result was not that specific. However, another subject complained that the information the top ranked pages provided was too general, "*I'm looking for more specific info, and not the general [information] that the top ranked page provide*." In the second prototype, the search result is presented in both thumbnail images and text. Most subjects preferred text to images and explained that it was mainly attributed to the limited screen size. Although most subjects complained that the web page thumbnails were too small, they could not deny the fact that such visual aid was somewhat useful, as one subject commented on the text/thumbnail tradeoff: *"I prefer text, but if I see a terrible website [on the thumbnail] I don't need to go there"*.

On searching for keywords: Several subjects reported difficulties in understanding the searching for keywords strategy, as complained by a subject: "*The keywords lead me out in tangents*". It might be an interface issue since both searching for similar web pages and searching for keywords strategies used a similar graphical design. Despite this, we also saw evidence of the effectiveness of the searching for keyword strategy. Ten subjects found some information that they were looking for in less than 5 steps (i.e. take a picture → search for keywords → select some keywords for Google → choose a webpage from the result returned by Google → browse and evaluate the information available on the web page). One subject suggested that he would *"... use keyword for fast searches and similar web pages for general information"*.

On searching by mobile MapQuest: Almost all subjects expressed the opinion that at first they thought such a map-based interface could be very helpful. But after using the map interface and the image-based approach, several commented that a map-based strategy might not be as helpful as it first seemed; in our tests users with the map-based interface often failed to find specific information about a location. But also, the small screen and the low resolution created many design tradeoffs. For instance, the mobile MapQuest was rejected by most subjects as too small and basically unreadable, as one subject commented: "*I have no idea what this small map is covering*".

In conclusion we found that simmilar web pages was the most intuitive interface but it fail in being able to find more specific information. The second strategy, keyword boosting, did a better job and provide more efficient searches but sometime failed due to non-relevant keywords and usability problems. Last, the specific map-based interfaces that we used in this test failed to provide useful information more than half of the time.

Table 1. Quantitative comparison of search strategies. Each row shows, under each search strategy, the number of attempts where the user found relevant information within 30 sec, or spent more than 30 sec, or did not find anything useful at all.

Search Strategy	<30 Sec	>30 Sec	Did not find
Web Pages	6	4	22
Extracted Keywords	10	12	10
MapQuest	4	10	18

7 Future Work

While we have demonstrated the feasibility of several CBIR techniques for our IDeixis system, it remains a topic of ongoing research to find the optimal algorithm to support searching the web with images acquired by camera-equipped mobile devices. We are exploring machine learning techniques which can adaptively find the best distance metric in a particular landmark domain. Additional study is needed to evaluate the performance of mobile CBIR with a range of different camera systems, and under a wider range of weather conditions (e.g., we did not consider snow, or darkness, in our studies to date.)

Currently the cameras on our prototypes have a fixed focal length and there is no simple way to specify a very distant location that occupies a small field of view in the viewfinder. We plan to augment our interface with a digital zoom function to overcome this, or implement a bounding box selection tool and/or use camera phones with adjustable zoom lenses when they become available.

As discussed earlier, even with image-based search of location-based information, additional context will be needed for some specific searches. Keyboard entry of additional keywords is the most obvious option, but with equally obvious drawbacks. Allowing users to configure various search preferences can be another option. An appealing interface combination would be to consider keywords obtained via speech input at the same time an image-based deixis was being performed (e.g. “Show me a directory of this building!”).

In addition, it is fairly likely that mobile devices in the near future will incorporate both camera technology as well as GPS technology, and that geographic information of some form may become a common meta-data tag on certain web homepages. Given these, we could easily deploy a hybrid system which would restrict image matching to pages that refer to a limited geographic region, dramatically reducing the search complexity of the image matching stage and presumably improving performance.

Finally, further study is very much needed to evaluate the usability of the overall method in various contexts beyond tourist and travel information browsing, including how to best present browsing interfaces for particular search tasks.

8 Conclusion

We have proposed an image-based paradigm for location-aware computing in which users select a desired place by taking a photograph of the location. Relevant web pages containing matching images are found using content-based image retrieval with the web or other large database. In contrast to conventional approaches to location detection, our method can refer to distant locations and does not require any physical infrastructure beyond mobile internet service.

We explored two interface paradigms, one based on thumbnail browsing of directly matched pages and one based on bootstrapped keyword search. In the latter approach keywords are extracted from matched images and used for a second-stage search, so that even if the best page for a particular location contains no image of the location, it still can be found using our system. We evaluated our prototype systems and found that directly matching simmilar web pages provide an intuitive way of searching for information that involves much less interaction than conventional interfaces, but can also yield in search results that sometimes are too general. On the contrary is the keyword based interface more cognitive demanding but can find more specific information faster.

References

1. Arasu A, Searching the Web. *ACM Transactions on Internet Technology*, 1(1), Aug. 2001, Pages 2–43.
2. Aslandogan, Y.A., and Yu, C. Multiple evidence combination in image retrieval: Diogenes searches for people on the web. In *Proc. of ACM SIGIR 2000*, 88-95.
3. Built-in camera mobile phones increasingly used for business purposes. *Nikkei Net Business*, Dec. 9 2002.
4. Cheverst, K., Davies, N., Mitchell, K., and Efstratiou, C. Developing a context-aware electronic tourist guide: some issues and experiences. *Proc. of CHI 2000*, 17-24.
5. FaceIT, www.visionics.com.
6. Gausemeier I., and Bruederlin B. Development of a real time image based object recognition method for mobile AR-devices. In *Proc. of the 2nd AFRIGRAPH*, 2003, 133-139.
7. Google, Google's PigeonRank, http://www.google.com/
8. Jansen, B. J., Spink, A., Bateman, J., and Saracevic, T., Real life information retrieval: A study of user queries on the Web. *ACM SIGIR Forum 32*, 1 (1998), 5-17.
9. Kaasinen E., User needs for location-aware mobile services, *Personal and Ubiquitous Computing* (2003) 7: 70–79.
10. Kosecka, J., Zhou, L., Barber, P. and Duric, Z. Qualitative Image-Based Localization in Indoors Environments, *CVPR 2003*, (To Appear).
11. Marmasse N. and Schmandt C., Location-aware information delivery with commotion. *HUC 2000 Proceedings*, 157-171.
12. Montemerlo, M., Thrun, S., Koller, D. and Wegbreit, B. FastSLAM: A factored solution to the simultaneous localization and mapping problem, In *Proc. of the AAAI National Conference on Artificial Intelligence*, 2002.
13. Niblack W., R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos and G. Taubin. The QBIC project, *SPIE Storage and Retrieval for Image and Video Databases*, 1993, 173-187.
14. Nokia, Next Generation Mobile Browsing, http://www.nokia.com/

15. Oliva, A., and Torralba, A. Modeling the shape of the scene: a holistic representation. *Intern. Journal of Computer Vision*, 42(3), 147-175.
16. Pilu, M. and Pollard, S. A light-weight text image processing method for handheld embedded cameras. *Tech. Report*, HP Laboratories, March 2002.
17. Schiele, B., Jebara, T., and Oliver, N. Sensory-augmented computing: wearing the museum's guide. *Micro. IEEE*, 21(3), 2001, 44-52.
18. Smith J. and Chang S., An Image and Video Search Engine for the World-Wide Web, *Symposium on Electronic Imaging*, IS&T/SPIE, Feb. 1997.
19. Smith, J.R., and Chang, S.F. Searching for images and videos on the WWW, *Multimedia Systems*, 1995, 3-14.
20. Swain, M., & Ballard D., Indexing via color histograms. *Intern. Journal of Computer Vision*, 1990, 390-393.
21. Symbian OS – The Mobile Operating System, http://www.symbian.com/.
22. Starner, T., Schiele, B. and Pentland, A. Visual Contextual Awareness in Wearable Computing. In *Proc. of ISWC*, 1998, 50-57.
23. Yeh T., Tollmar K. and Darrell T. Searching the Web with Mobile Images for Location Recognition, *In proc of CVPR 2004,* (accepted for publication).

The Use of Landmarks in Pedestrian Navigation Instructions and the Effects of Context

Tracy Ross, Andrew May, and Simon Thompson

Ergonomics & Safety Research Institute, Loughborough University, Holywell Building, Holywell Way, Loughborough, Leics, LE11 2SE, UK
{t.ross, a.j.may}@lboro.ac.uk

Abstract. The beneficial effects of using landmarks in vehicle navigation systems (improved user confidence and navigation performance) have been well-studied and proven. The study reported here aimed to investigate the effects of adding landmark information to basic *pedestrian* navigation instructions (i.e. those which included distance to turn and street name only). The study found that the results replicate that for vehicle navigation systems. User confidence was raised to a consistently high level as a result of landmark inclusion and errors were greatly reduced. The results also indicate the types of manoeuvre that should benefit most from the inclusion of landmarks.

1 Introduction

Navigating in an unfamiliar location can be aided by the provision of navigable database information via intelligent route guidance devices. Route navigation technology has, until now, been developed primarily for in-vehicle use. More recently, the databases and routing intelligence have been used to develop pedestrian navigation applications typically available via mobile phones or PDAs (Personal Digital Assistants). Information on the user's position is gathered via manual entry of start and end points, accurate technology such as the Global Positioning Systems (GPS) and, for mobile phone-based services, cell positioning

Research conducted on the navigation requirements of both drivers [1], [2] and pedestrians [3] has indicated that landmark information is of vital importance. Many of the currently available systems for navigation contain little or no information of this type. More detailed research into the use of landmarks for vehicle navigation has indicated that the main impact of reference to such information is an increase in the user's confidence in the navigation system and the reduction of errors [4]. The impact of landmark information on users of pedestrian navigation has not been assessed in the same way. This paper provides new data to determine this effect.

The specific aims of this study are to investigate whether the addition of landmarks to pedestrian navigation instructions (a) increases user confidence and (b) improves navigation performance (i.e. number of correctly completed manoeuvres), and whether any effects occur consistently for all manoeuvres.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 300–304, 2004.

2 Method

The study consisted of a pedestrian trial involving 40 participants using text-based navigation information (provided on flip cards) to walk an unfamiliar route in an unfamiliar urban environment (all participants reported their knowledge of the area prior to the trial as 'poor'). The participants were divided into two matched groups (based on age and gender) with a 50/50 split of males and females in each group. Participants' age range was 20-41 years with an average age of 24 years in each group.

One group was provided with 'basic' navigation instructions and the other group 'enhanced' navigation instructions. The basic navigation instructions were based on those provided by a current, commercial pedestrian navigation provider. An example of a basic instruction was 'After 0.1 miles turn right and continue onto Church Gate'. The enhanced instruction included landmarks and a corresponding example is 'After 0.1 miles turn right after the **church** and continue onto Church Gate'.

The route consisted of 18 manoeuvres (i.e. points at which the user had to follow an instruction). At ten of these manoeuvres it was possible to enhance the instruction by providing a landmark. For the others no landmark was available and road layout information was provided to enhance the instructions to a similar level. This paper concentrates on analysis of the 'landmark-enhanced' manoeuvres only.

For each manoeuvre, participants were stopped at appropriate points (termed 'individual confidence rating' points) and asked to give a rating of their confidence in the next manoeuvre they had to make (based on the navigation instruction they had received). These points along the route were chosen because there was some 'change' at that point, e.g. the pedestrian was turning a corner and the road scene changed or a road sign became visible). The rating scale was: 1 – very unconfident; 2 – unconfident; 3 - neither confident nor unconfident; 4 – confident; 5 – very confident. If a navigational error was made, the participant was directed back onto the correct route. The confidence data for manoeuvres that were incorrect were excluded from the analysis.

3 Results

For each manoeuvre, a mean rating of confidence was calculated (across the confidence rating points for that manoeuvre). Figure 1 shows the mean confidence ratings for each manoeuvre depending on whether the participants were provided with basic or landmark-enhanced instructions. For the basic instructions, ratings were very variable and fell between 2.4 and 4.4. For the enhanced instructions the range was smaller and consistently high: 4 - 4.8, i.e. always above the 'confident' level.

Each individual confidence rating (there could be up to 5 for each manoeuvre), was subjected to a Mann Whitney statistical test to compare the difference between the confidence ratings for the basic and enhanced set of instructions. Of the total 36 points (over 10 manoeuvres) 29 showed a significant difference ($p<0.05$) with all points showing higher confidence for the enhanced instructions.

Figure 1 also shows that for some manoeuvres, the confidence benefit of adding landmark information was greater than for others. Manoeuvres 2, 3, 5 and 8 all had an increase in confidence of more than 1 which corresponds to one step up the confidence rating scale.

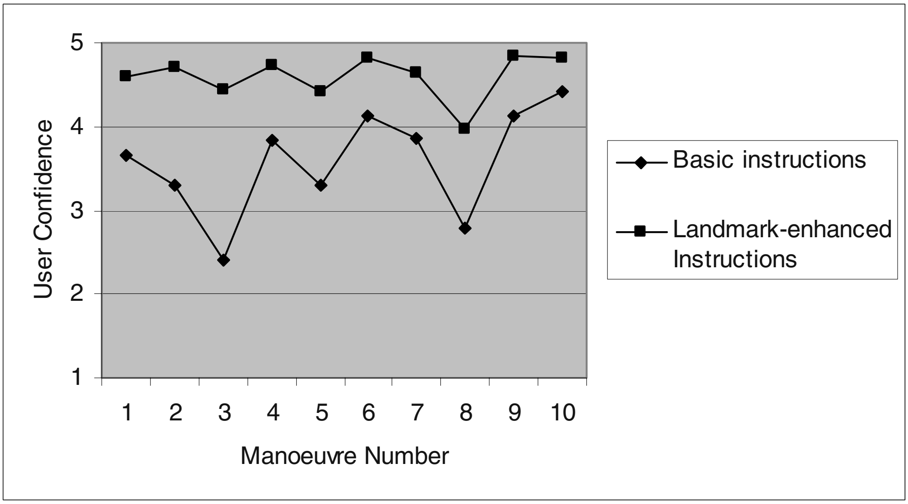


Fig. 1. Mean confidence ratings per manoeuvre with basic or landmark-enhanced instructions

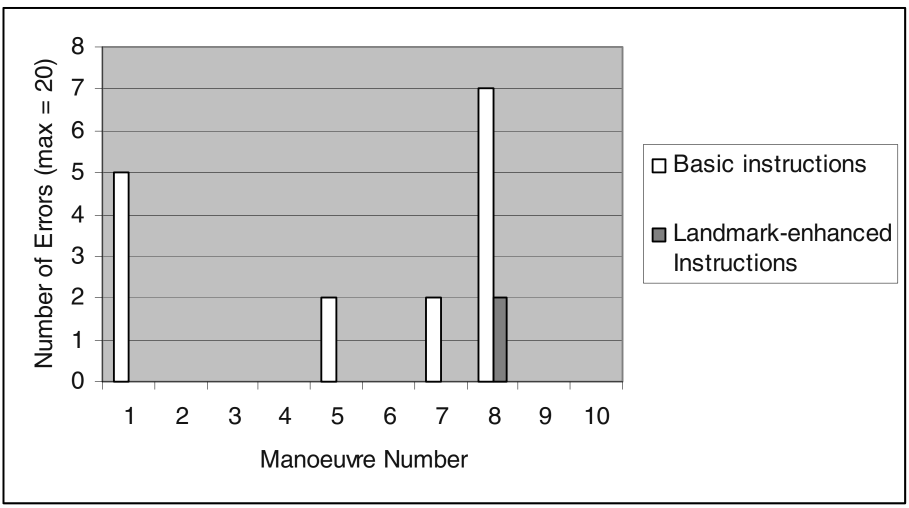


Fig. 2. Number of participants making errors with basic or landmark-enhanced instructions

Figure 2 shows the number of participants making errors at each manoeuvre. The total number of errors made across the ten manoeuvres was 16 (8%) for the basic

instructions and 2 (1%) for the landmark-enhanced instructions. Manoeuvres 1, 5, 7 and 8 appeared the most problematic in navigational terms.

4 Discussion

The inclusion of landmarks within the pedestrian navigation instructions increased user confidence and reduced or eliminated navigational errors in all cases. One main effect on confidence levels was the *consistency* of the confidence ratings. This is particularly beneficial to the user's perceptions of the system. If a system, in most cases, supports the user but there is one situation where this is not the case, the user's attitudes towards the system overall may be damaged.

Participants' comments throughout the route indicated that the reason behind this general improvement in confidence was that the landmarks assisted users in identifying the precise location of the manoeuvre at a greater distance back from the manoeuvre. This was because the users were able to see the object in question prior to seeing a street name (one of the main components of the basic instructions). In the environment in question (as in most of the UK), street names are often only visible when very close and sometimes are not present at all, or only on one side of the road. The basic instructions also relied heavily on the judgement of distance, a skill which varies widely amongst users. These findings are consistent with those in studies relating to the use of landmarks for *vehicle* navigation [1], [2], [4].

Another related finding was that due to this raising of confidence to a consistent level, navigational confidence at some manoeuvres (compared with the majority of manoeuvres) was increased by a much greater amount (i.e. those manoeuvres where navigation had been particularly problematic when using the basic instructions). Three of these manoeuvres (Numbers 3, 5 and 8) shared common features of being in pedestrianised areas where there were many choices of direction. Some of these were named streets (usually the intended route to take) and some were private entrances or alleyways. However, this differentiation was not significant to the pedestrian as, *visually*, the difference did not help to indicate the most likely direction. In cases such as these, the addition of a visual landmark is particularly beneficial. The one other problematic manoeuvre (Number 2) did not share these features. From assessment of the route, it is likely that the problems with this manoeuvre stemmed from the fact that it was a minor (narrow) road, leading from the main (wide) route. Past research has indicated transfer from a wide route to a narrow one is one of the navigation scenarios where extra support is usually needed [5].

The reduction in navigational errors was also consistent across all manoeuvres when the enhanced instructions were used. For manoeuvres 5, 8 and 12, the navigational errors experienced when using the basic set of instructions can probably be explained in the same way as the low confidence levels. These were both manoeuvres in pedestrianised areas with many (equally likely) choices and low (or no) visibility of street name signs. Manoeuvre 1 also caused a high level of errors (for the basic instruction set). This was the first manoeuvre to be made, the street sign was not very visible and there were two equally likely directions that could be taken.

5 Conclusions

The study showed that landmarks can be beneficial to pedestrian navigation in the same way as they have been proven to be valuable in vehicle navigation. The main benefits are: an increase in user confidence, an improvement in navigation performance (reduction in errors) and the provision of a more 'consistent' user experience (confidence across all manoeuvres had high variability when relying on street names and distance judgement and high consistency when the instructions were enhanced with landmarks).

The results also showed that some types of pedestrian navigation manoeuvre can be more problematic than others (i.e. where the intended direction is less obvious due to the user having many, equally likely, choices of direction such as in a pedestrianised area or where the change is from a major to minor road.

If technologies such as PDAs and mobile telephones are to provide pedestrian navigation applications, incorporating landmarks within visual (or verbal) instructions is likely to enhance the user experience, increase the uptake of such services and encourage continued use.

References

1. Streeter, L.A., Vitello, D.: A profile of drivers' map-reading abilities. Human Factors, 28(2) (1986) 223-239
2. Burnett, G.: 'Turn right at the traffic lights': the requirements for landmarks in vehicle navigation systems. Journal of Navigation, 53(3) (2000) 499-510
3. May, A.J., Ross, T., Bayer, S.H., Tarkianen, M.J.: Pedestrian navigation aids: information requirements and design implications. Personal & Ubiquitous Computing, 7(6) (2003) 331-338
4. Ross, T., May, A.J., Duffield, J.M., Paszkowicz, S.: Driver reaction to navigation instructions incorporating good and poor landmarks. EPSRC LINK Inland Surface Transport programme, REGIONAL Project Deliverable 4, Ergonomics & Safety Research Institute, Loughborough University, UK (2002)
5. Sugiyama, H., Hasegawa, T., Doi, M: A pedestrian navigation system based on a navigation demand model. Toshiba Research and Development Centre, Japan (2001)

A Paper Prototype Study of the Interface for a Children's Collaborative Handheld Learning Application

Jason T. Black[1], Lois W. Hawkes[1], Ketly Jean-Pierre[2], Isaac Johnson, II[2], and Marvin Lee[2]

[1] Computer Science Department
Florida State University
Tallahassee, Florida 32306
{ black, hawkes }@cs.fsu.edu
[2] Computer Information Sciences Department
Florida A&M University
Tallahassee, Florida 32307
{ kjeanpie, ijohnson, malee }@cis.famu.edu

Abstract. The use of handheld computers in the classroom environment is an area generating much interest among researchers. The questions of how to manage power and graphics while handling small screen space remain issues to be examined. In this study, a paper prototype test of the collaborative interface features was implemented using five user tasks: logging in, reading a text, answering text related questions, chatting, and entering data into a personal workbook. The test results illustrated the need for clear instructions and menu options for younger users, and that speech and written input were preferred over other methods. The feedback obtained from the test will drive the development of the actual system in future studies.

1 Introduction

Handheld computers have achieved a significant amount of attention in recent years as an alternative to the bulky, stationary, cable-dependent desktop personal computer [3], due to small size, low power consumption, portability, and similarity to game devices with which students are already familiar [3,8]. Unfortunately, the limitation presented by the smaller screen size directly affects what can be accomplished in the way of user interface development for such devices [1] [3] [8].

In this paper, we discuss the development of a paper prototype test of the interface for a learning application for the personal digital assistant (PDAs), and the implementation of the test using potential end-users. The goal was to obtain a clear understanding of not only what features aid in the development of learning, but also what strategies can be employed to facilitate as smoothly as possible the overall learning process.

2 System Overview

The learning system modeled in this study was centered on a reading comprehension application. Reading comprehension is an area of much concern, as educators and re-

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 305–309, 2004.

searchers alike are examining ways to improve the method by and rate at which students acquire reading comprehension skills [10]. Many studies have focused on reading, but few have incorporated handheld computers into this effort [9].
Students will be working in pairs while progressing through collaborative reading lessons. All of the activities will be done on the student's PDA, with lessons being transmitted wirelessly from the server to the PDA as requested, and student progression data being stored on the system server. This research effort explored the effective development of these interface designs in the handheld environment, and paper prototyping was used to obtain answers on how to best accomplish this feat.

3 Paper Prototype Usability Study

The paper prototype study was conducted with the assistance of a local after-school program, with students in the range of grades 2 to grade 4. The study had 5 participants, based on recommendations from Snyder's research on the acceptable number of test subjects for a paper prototype study [7], as well as Nielson and Landauer [4]. Subject #1 was a fourth grader who was a good and constant reader, and was the only subject that had limited familiarity Subject #2 was a third grader who had some experience with comput t a strong reader. Subject #3 was a second grader with very little com e, but was a strong reader. Subject #4 was a third grader who had very experience and was also not a very strong reader. Subject #5 was a fou had computer experience, but was having trouble reading at grade leve dents had some exposure to the basic features of a computer, though er previously seen or used a PDA prior to the testing period.
The study was conducted using mc rface done on paper. This involved a small cardboard representation of hp iPAQ® 5555 PDA [Fig.1]. The screens were developed using a so package instead of the hand-drawn designs seen in most paper prototy e to the age and experience of the students being tested. [6].

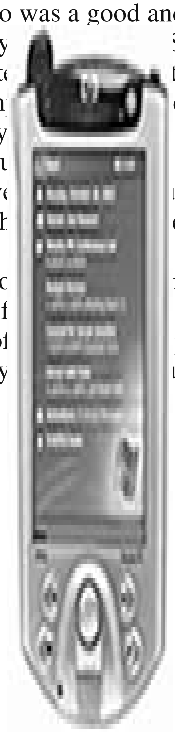


Fig. 1. Picture of actual hp iPAQ® 5555 handheld computer

The paper prototype testing centered around the completion of five tasks most signification to the functionality of the application: logging into the system; reading through textual paragraphs; choosing answers based on the text passage read; utilizing the

chat/message sending facility; and entering data into a personal notebook feature. Logging into the system involved the subject being asked to enter his or her user name and password either by using a virtual keyboard or by writing, and selection an icon to represent them.. The virtual keyboard was designed as a small cutout of a set of keys that could be placed onto the screen using removable tape. The writing space was a square cutout of a piece of transparency paper that could also be taped on the screen, all recommendations by Snyder [7].
Reading the texts involved the subject being asked to read a short age-appropriate story presented on a page of text with a "Back" and/or "Next" button for turning pages. The answering questions task presented the user with a short question and a series of possible answers represented as radio buttons which were implemented using removable tape over the buttons drawn on the screen [7] [Fig 2]. The messaging task required each subject to indicate he or she needed assistance from the instructor by either keypad, written or voice input (via microphone). The data entry task required the subject to write his or her name in the personal journal with the same input guidelines as in other tasks.

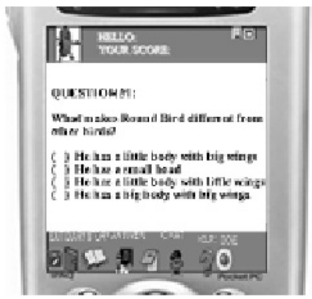


Fig. 2. Screen shot of Question Answer screen.

At the completion of each task, each subject was asked to talk about his or her experience and their comments were used to modify the test to incorporate the recommendations prior to subsequent tests. Subjects were tested in 30-minute sessions, with 2 tests conducted on a given day, as recommended by Snyder [7].

4 Results and Discussion

The results of the testing were very enlightening. During the logging-in task, 3 of the 5 subjects preferred writing their name to using the virtual keypad. Of the 2 subjects that did choose to pick their letters using the keypad, one subject (Subject #4) had trouble navigating the keypad, and began pressing buttons on the bottom of the PDA instead of the buttons on the keypad. It was discovered that he was familiar with GameBoy™ handheld computer games, which use the directional keypad for manipulation of all applications. All subjects except Subject #1 had trouble finding the icon during the player selection phase. All subjects had no trouble finding and clicking the "Done" button.

All subjects were able to manipulate the reading task with no problems or difficulties, indicating that screen real estate was not a significant issue here [1, 2, 3]. During the question-answer task, subjects were able to move freely from page to page, and the positioning and size of the text in the passage was adequately displayed. The chat/messaging task perhaps provided the most valuable feedback, with all except one subject choosing to speak their message instead of the other input features (Subject #4 chose to use the keypad in all writing tasks). However, two of the subjects were confused when faced with the submenu that appears with the speak feature, which asks them to click on the microphone to begin speaking and to click send to transmit the message. But once this was explained, the subjects were able to complete the task.
Finally, the data entry task was easily understood and manipulated by all subjects. The choice of writing in the diary as opposed to using the keypad was selected by 3 of the 5 subjects.
Below is a listing of those errors and a discussion of how significant those errors may be in overall efficiency of the system.

1) *Confusion on wording of screens*: some of the wording of the instructions on some screens was confusing to the subjects, possibly too advanced for this level of student. Modification of these screens yielded a marked improvement in later sessions [7].
2) *More instructions before tasks*: it was recommended by several of the subjects that an introduction screen be presented prior to the beginning of each task, possibly as a feature that can be toggled off for more advanced users.
3) *A "Clear" feature on writing tasks*: it was suggested by subjects that a clearing feature on the writing area would be very useful in making the writing feature easier to use. It is possible that this mechanism is a built-in feature of the iPAQ® system; this will be answered during development.
4) *The speaking instead of writing of messages*: most subjects selected this feature as a means of transmitting data between users. The iPAQ® 5555 does indeed provide such capability.
5) *Simplification of the speaking submenu*: It was suggested that the buttons that are not useful during this phase be grayed or toggled off to avoid confusion.

It is significant to note that those changes resulting from these recommendations yielded improved performance during the second and third day of testing. This incremental development and modification will hopefully result in us having reached an optimum design for future implementation of the system.

5 Conclusions and Future Work

The data obtained from the paper prototype testing was extremely helpful in assisting developers in identifying potential problem areas in our user interface design. It is important to note that these issues underscore the difficulty in developing applications for younger users – adult researchers often have difficulty understanding and estimating what younger users would like to see in an interface. This illustrates why design techniques such as paper prototyping are paramount in gleaning concise and efficient

specifications from such a target group. It is also important to note our intent here was to obtain a measure of how to structure the features of the interface, not to report on the results of an actual empirical test of the software being implemented in the classroom. It is apparent that we have only scratched the surface in examining what can be done and what is desirable in handheld interfaces for children. But it is clear that the intersection of collaboration and technology in such a user-driven fashion can lead to a much more robust learning environment.

Acknowledgements. We would like to thank the members of our Educational Technology group at Florida State University; the student workers at Florida A&M University, the CIS department at Florida A&M University, and the participants in this study and the staff at the New Beginnings after-school. This work is funded by the Florida A&M University Faculty Research Pilot Program Grant [G11 HD41829-02], sponsored by the National Science Foundation.

References

[1] Bederson, B., Clamage, A., Czerwinski, M., Robertson, G. A Fisheye Calendar Interface for PDAs: Providing Overviews for Small Displays. *In Proceedings of CHI 2003*, ACM Press, pp. 618 – 619.

[2] Genise, P., *Paper Prototyping: A Tutorial*, version 1.2., 2002, http://www.cs.utexas.edu/users/almstrum/cs370/elvisino/paperProto.html.

[3] Mandryk, R.L., Inkpen, K.M., Bilezikjian, M., Klemmer, S.R., and Landay, J.A (2001). Supporting children's collaboration across handheld computers. In Extended Abstracts of *CHI, Conference on Human Factors in Computing Systems*. Seattle, USA, April 2001.

[4] Nielsen, J., and T.K. Landauer, A mathematical model of the finding of usability problems. In Proceedings of Conference on Human Factors in Computing Systems: INTERCHI '93 (Amsterdam), pp. 206-213. New York: ACM Press.

[5] Online Computer Library Center, Prototyping: How we do it: Paper Prototyping, 2002, http://ww.oclc.org/usability/prototyping/oclc.htm.

[6] Rettig, M. Prototyping for tiny fingers. *Communications of the ACM*, April, 1994, Vol.37, No.4.

[7] Snyder, Carolyn. Paper Prototyping: The Fast and Easy Way to Design and Refine *User Interfaces.* Morgan Kaufmann Publishers: San Francisco, 2003.

[8] Yee, K. Peephole Displays: Pen Interaction on Spatially Aware Handheld Computers. *In Paper/Demos: Interaction Techniques for Handheld Computers, CHI Letters 2003*, 5(1), pp. 1 – 8.

[9] Computers and Classrooms: The Status of Technology in U.S. Schools (1997). Policy Information Center. Educational Testing Service. Princeton, NJ.

[10] National Center of Educational Statistics. (2001). National Assessment of Education Progress (NAEP) 2000 reading report card for the nation and states. Washington, D.C.:

Ubibus: Ubiquitous Computing to Help Blind People in Public Transport

Michel Banâtre, Paul Couderc, Julien Pauty, and Mathieu Becus

INRIA
Campus Universitaire de Beaulieu
F-35042 RENNES
{michel.banatre, paul.couderc, julien.pauty, mathieu.becus}@irisa.fr

Abstract. Ubibus is an application designed to help blind or visually impaired people to take public transport. The application allows the user to request in advance the bus of his choice to stop, and to be notified when the right bus has arrived. The user may use either a PDA (equipped with a WLAN interface) or a Bluetooth mobile phone.
The system is designed to be integrated discretely in the bus service via ubiquitous computing principles. It tries to minimize both the amount of required changes in the service operation, and explicit interactions with the mobile device. This is done by augmenting real-life interactions with data processing, through a programming paradigm called *spatial programming*.

1 Introduction

With the development of mobile devices, we can now use computers in various conditions and places. Usually, when the user requests a service from his mobile terminal, he is already involved in an activity, like shopping, visiting a museum or having a meeting. In such situations, his attention is a scarce resource: he has a limited capacity to interact with the computer, and therefore the applications must deliver services with minimal interactions.

Our approach to reduce the interactions between the human and the computer is to annotate physical interactions with computing instructions. Consider for an example a shopping cart. The two possible physical interactions are: adding a product inside the shopping cart, or withdrawing one. If we can detect the physical interactions then we can associate them to computing operations. For the shopping cart, when it detects the first interaction then the price of the product is added to the total price of the shopping cart. Conversely, the shopping cart subtracts the price of the product when it is withdrawn.

The main purpose of this approach is to support the creation of applications that are directly driven by physical interactions, which naturally reduce the need of interactions with the device. We call it *spatial programming*. In this article we focus on an application, *UbiBus*, which relies on spatial programming. The main objective of this application is to help blind or partially blind people to take the public transport. The next part presents UbiBus scenario. Section 3 is

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 310–314, 2004.

dedicated to spatial programming and its benefits from an HCI point of view. Before concluding, we present in section 4 some related works.

2 UbiBus

2.1 Context

Bus transportation systems work in a similar way in different countries. To take a bus we just have to go to the closest bus stop and make a sign to the bus driver when the bus arrives. But this simple task is complicated for a blind or visually impaired person. In this article we call him Peter. Consider the following situations:

– Peter is alone at a bus stop where several busses stop. He cannot see the busses, so he cannot signal the bus to stop and his bus does not stop.
– Peter is alone, but this time he can see the busses but cannot read the line numbers. In this situation he signals every bus to stop.
– Peter is not alone at the bus stop and ask for help to another person

In every case Peter cannot read the bus schedule so he does not know how much time he has to wait.

From these examples we can see that Peter's handicap prevents him from using public transport easily. Our objective is to propose to Peter the *UbiBus* application, which helps him to take the public transport. This application should be easy to adopt, and should not disturb the service for other users and bus drivers.

2.2 A Typical Scenario

We will show how UbiBus works by considering the typical usage scenario. Three types of entities will interact: the bus riders (like Peter), the bus stop, and the bus.

Asking the Bus to Stop. Peter has a mobile phone equipped with a short range communication interface such as Bluetooth or WiFi. Peter interacts with UbiBus via speech recognition. The only thing he has to do is to say the bus route number Once Peter has said which bus he wants to take, he walks toward the bus stop. When he is close enough from the bus stop, his phone notifies him with the estimated time to wait (6 minutes for instance), received from the bus stop.

Stopping the Bus. Peter is still waiting for his bus. Another bus is approaching, but Peter cannot see it. However, inside the bus, the driver notices a flashing "stop request" message displayed on the screen of device installed on the dashboard The driver stops the bus, opens the door and Peter is notified by a vocal message that *his* bus has arrived.

3 Spatial Programming

Spatial programming [1,2] is a principle which consists in expressing a program in terms of spatial interactions between physical objects representing data. Spatial programming is based on an analogy in which the physical space is considered as a data store, and physical movements are considered as a way of *addressing information*. In spatial programming, reading a data item x means *physically moving* near the object (or zone) representing x, *or* waiting for an object representing data x to come close to the reader. SPREAD [2] is the spatial programming model and an execution environment supporting Ubibus. An important feature of this system is that each physical entity participating in the system is *autonomous*, which means that it has its own computing device and communication capability with nearby devices, operating in peer to peer mode.

3.1 An Overview of UbiBus Implementation with SPREAD

SPREAD consists in associating data items to objects and synchronizing data processing on physical encounters of objects. Synchronization is based on both data matching and physical proximity, meaning that a program looking for a data will remain blocked until the relevant data is visible in range.

In Ubibus, we have to consider three entities: the bus rider (Peter), the bus stop, and the bus itself. Each of these entities will run a spatial program, supported by SPREAD. The main interactions that control the spatial programs of UbiBus are: first, the encounter of Peter and the bus stop, which activates a stop request for the relevant bus. And second, the encounter of the bus with the bus stop, which signals the bus driver to stop. Other interactions are supported by the system, such as next stop announcements for the bus passengers, spontaneous display of the local map as a passenger gets off the bus, contextual advertisements access linked to the paper advertisements on the bus shelters and so on. The same idea of "annotating" physical interactions with actual processing is used to support all these services, but are beyond the scope of this paper which present the system in the context of visually impaired users.

3.2 Spatial Programming and Human Computer Interactions

The main goal of spatial programming is to piggyback onto physical processes that already exist in today's services or tasks in order to enhance them. One goal is to reduce as much as possible explicit interactions between the user and the computer. Spatial programming offers to the application developer a framework in which data processing is directly expressed in terms of interactions between physical objects, promoting spontaneous operation of the software.

In our opinion, this aspect is especially important for an easy adoption of such enhanced services, enabled by "embedded intelligence". People are focused on their real-world tasks, and each explicit interaction with a computer disturbs the user from his activity. Context-awareness is one way to guess the user's situation

and to try to reduce inputs required from the user. Usually, a context-aware application uses environment sensing methods (like GPS positioning, sensors etc.) to automatically adapt the behavior of an application. The spatial programming approach goes further by avoiding the use of an intermediate representation of the context: the data structures, data processing, and program logic must be mapped directly on physical processes or interactions, making applications control non intrusive.

However, in some cases asking the user for input cannot be avoided, because we cannot always identify an existing interaction in the service on which we could map an operation of the program. In these cases, we tried to use the simplest user input options available. The bus rider is the entity which has the most complex user interface issues in UbiBus: it requires an explicit input for the number of the bus to stop, and also offers access to additional features such as relevant information for the area (maps, advertisements...).

User Terminal. The bus rider can use either a PDA or a Java-enabled mobile phone. An important lesson learned of our experimentations with Ubibus is that there is a resistance of the user against unfamiliar devices such as PDA, much stronger than what we had expected. An interesting point to note is that the user very easily learns how to use the application on a familiar device (the cell phone) even though some interactions are *less easy* than with the touch screen interface of the PDA.

Stop Request Selection. Selection of the bus to stop is achieved by speech recognition (IBM ViaVoice). Time to wait for the requested bus is announced by a voice message as soon as the user's device is in the range of the bus stop. The display area of the device is used to display in a cycling way the set of information available at this bus stop. Clicking (if the device has a touch screen) or pressing the selection key of the device (Yes on a mobile phone) allows the user to open the information.

4 Related Works

Helping visually impaired people by providing enhanced perception through smart objects/space annotation has already been proposed, for example in [3,4]. However, these systems do not propose a general programming model for *annotating* physical interactions with program code, unlike spatial programming.

Like UbiBus, Bus Catcher [5] is also a public transport helper application. Essentially, Bus Catcher displays on the users PDA the accurate and timely timetables for all bus routes. An important difference with UbiBus is that Bus Catcher relies on explicit user-computer interactions, while in UbiBus the application control is implicitly mapped onto the real life interactions of the existing service as much as possible.

Another interesting project is the Human Pacman [6], which proposes an outdoor Pacman game involving physical interactions of moving people. Unlike

Ubibus, this system is based on a centralized information server. With spatial programming, the Pacworld would be built by physically disposing the game items (cookies etc.) in the game field, each one including a small embedded computer.

5 Conclusion and Future Work

The Ubibus application presented in this paper shows how *existing* services can be enhanced by ubiquitous computing in a non intrusive manner. This is done by annotating interactions already existing in the physical world with computing instructions. This approach is especially effective in the case of Ubibus where the existing service is highly dependent on interactions involving mobile entities.

The Ubibus application shows several enhancements in the service: the main one is helping visually impaired people to catch their bus. The system also helps other people by providing the waiting time (indicating whether to rush or not), contextual advertisements and information linked to the bus stop, such as a local map or a movie trailer.

Our current works involve the development of new applications based on proximate interactions. We are working on enhanced spatial programming model to support a wider range of applications.

References

1. Couderc, P.: Mobilit contextuelle dans les syt mes d'information. PhD thesis, Universit de Rennes 1 (2001)
2. Couderc, P., Banâtre, M.: Ambient Computing Applications: an Experience with the SPREAD Approach. In: 36th Annual Hawaii International Conference on System Sciences (HICSS'03), Big Island, Hawaii (2003)
3. Goto, K., Kambayashi, Y.: A New Passenger Support System for Public Transport using Mobile Database Access. In: Very Large Data Bases (VLDB'02). (2002)
4. Coroama, V., Röthenbacher, F.: The Chatty Environment - Providing Everyday Independence to the Visually Impaired. In: Workshop on Ubiquitous Computing for Pervasive Healthcare Applications at Ubicomp 2003, Seattle, Washington (2003)
5. Bertolotto, M., O'Hare, G., Strahan, R., Brophy, A., Martin, A., McLoughlin, E.: Bus Catcher: a Context Sensitive Prototype System for Public Transportation Users. In: Third International Conference on Web Information Systems Engineering (Workshops) (WISEw'02). (2002)
6. Cheok, A.D., Wan, F.S., Goh, K.H., Yang, X., Liu, W., Farzbiz, F., Li, Y.: Human Pacman: A Mobile Entertainment System with Ubiquitous Computing and Tangible Interaction over a Wide Outdoor Area. In: Mobile HCI'03. (2003) 209–223

Improving the Effectiveness of Mobile Application Design: User-Pairs Testing by Non-professionals

Titti Kallio and Aki Kekäläinen

TeliaSonera Finland, P.O. Box 585, FIN-00051 Sonera, Finland
{titti.kallio, aki.kekalainen}@teliasonera.com

Abstract. The nature of mobile applications requires a fast and inexpensive design process. The development phase is short because the life cycle of an application is limited, mobile technology is developing rapidly, and the competition is heavy. Existing design methods are time-consuming and require expertise (e.g. Contextual Design). We suggest a design approach where focus groups are followed by usability tests in pairs carried out by non-professional moderators. With this approach CHI departments can benefit from market research resources, and improve collaboration with marketing people. We evaluated this approach with a case called News Client. The findings show that in paired-user tests near half of the usability problems were found compared to individual usability testing. The results are not too profound but enough for industry needs. Another interesting point is that our findings do not support the earlier reported results according to which the interaction between two participants can bring out more input than a single participant thinking aloud.

1 Introduction

In the mobile industry there are more and more mobile applications designed and launched by different parties. The application design process is usually rapid; there is a need to design usable applications quickly and cheaply, because:

- Applications are ‘small’ or simple in the sense that they usually cover a user’s need to accomplish a certain single task (e.g. checking a bank account).
- A simple application often means short code (limited resources needed for programming).
- Time from the idea to the launch is limited (also due to the heavy competition).
- The product’s lifecycle is short (1-12 months) due to the development tempo in the field (technologies, networks, and devices are developing rapidly).
- Resources are limited; e.g. CHI experts in the field are still few.
- Also many applications are developed by small companies with even fewer or no CHI experts at all.

Also the possibilities in the field are expanding rapidly. In a few years manufacturers have launched colour handsets and colour WAP, XHTML for mobile browsers, Symbian Operating System with native clients running on Symbian, and Java support. The technology in the field is developing more rapidly than usage habits, so the

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 315–319, 2004.

mobile industry wants to launch new applications frequently to teach users new technology and service possibilities. The field is still very 'technology-push' oriented.

Most existing design methods used in large-scale information system development processes are unsuitable due to the lack of time and resources – there is a need for a design approach or toolkit, which suits this kind of 'small' and quick applications development. E.g. Contextual Design [1] and ethnographical approaches consume much time, financial resources, and require numerous experts.

The approach or design toolkit, which is needed, should be easy to learn and use, must not require many experts and should be based on familiar methods at least to some extent to avoid resistance in the organization. Also the approach may not be too detailed, because the need is just to cover basic usability issues – in practice there is little time for more.

Next we present some available methods and suggest a new approach based on existing ones. We also evaluate the new approach with a case called News Client.

2 Focus Groups and Usability Testing in Pairs

Marketing departments often use Focus Groups, a group discussion method, to find out which features to choose for a new product, how customers would use it, and how much it can cost. To organise a focus group takes some effort: you have to find people to participate, recruit them, organise the setting and give participants compensation. We suggest that CHI departments can take advantage of all this and involve the participants of a focus group in a usability test, which occurs just after the focus group discussion is over. The following usability test session must be short (15-30 minutes) because participants have already discussed for 1-2 hours in the focus group session.

The topic of the focus group and the tasks in the usability test can concern the same service or area, but they can as well be totally different (if the scope is the same, then the preceding group discussion probably affects the usability test session afterwards). The idea is to benefit from focus group situations as they are often organised anyway, and CHI people often have some smallish usability problems, which would be good to test in a fast and quick way.

We also suggest other benefits, for example with our approach CHI and marketing people will work more closely together and therefore learn to understand each other's work better. Also the finances are an issue, because this is a great way to decrease costs. For example in Finland a focus group study including 2-3 groups (6-8 participants in a group) costs approximately the same as a usability test for 8-10 participants if these are ordered from a research company. So it saves money if you every now and then can skip an expensive separate usability test.

The following usability test part could happen in a traditional way, i.e. the 'de facto standard' thinking aloud protocol analysis based on the theoretical framework by Ericsson and Simon [3], but this would require several professional moderators to be

available at the same time to do the testing. For example for a focus group of 10 participants, we would need 10 moderators to be ready after the group session ends. That is why we decided to apply a testing in pairs approach (also called as co-discovery or co-participation [5, 8]).

The pair-tests do not require expertise in the sense that a single thinking aloud test does. In a traditional thinking aloud test a user is given tasks and asked to think aloud while trying to accomplish these tasks. The technique in question is quite demanding and the moderator should be an expert. In pair-testing the tasks are given to two users, who try to accomplish them working together. While trying to solve problems together these two test users speak out naturally and that is why the situation is quite easy for the moderator. Therefore we suggest that usability testing in pairs can be done by less experienced non-professional testers.

Other reasons supporting the use of pairs instead of individual usability tests include problems with the thinking aloud protocol according to D. Wildman [7]: the individual test situation can be hard for the test user (if one feels himself or herself 'stupid'), and often the test user needs psychological positive feedback to keep the thinking-aloud process going on (moderator needs to e.g. mumble encouragingly constantly), whereas with user-pairs partners naturally converse. For all these reasons we decided to evaluate the user-pairs testing carried out by non-professionals.

3 Case 'News Client'

TeliaSonera Finland has launched an application called News Client, which can be downloaded to a mobile phone with Nokia's Series 60 Symbian Operating System platform. With the News Client you can view news and the weather report; also, the news can be updated automatically or on demand.

The News Client was usability tested in three different ways with 'typical' customers as test users:

Test I: Traditional usability tests with single users; the thinking aloud approach; professional CHI experts as moderators; 6 separate test users. This is the more expensive way of testing. The results of the third test group were compared to the results of this test group.
Test II: Usability tests in pairs after a focus group session (the focus group discussion topic was not the News); *professional* CHI experts as moderators; 3 pairs (i.e. 6 test users). This step was included to get more comparison data for the results.
Test III: Usability tests in pairs after a focus group session (the focus group discussion topic was not the News); *non-professional* moderators; 3 pairs (i.e. 6 test users).

CHI experts instructed non-professional moderators shortly (10-15 minutes), and told them not to disturb the pair working together, but to concentrate on writing notes. Moderators were also warned not to 'teach' participants, but to let test users to try to

solve the problems by themselves. The moderators were a project manager, a concept designer and a programmer with no previous experience in usability testing.

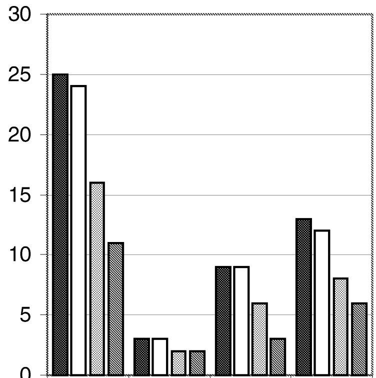
The same 13 tasks were used in every session, starting with the task to download the application from the WAP portal, to open it, to search some basic information (a certain news item, weather etc.), and to remove the application from the handset.

After the tests two usability professionals analysed the findings of every test group based on moderators' written notes. The findings were grouped in three categories (figure 1). After the categorisation the analysis of the three test groups was done in a few hours.

4 Results

In our tests different sessions revealed a different amount, but mostly the same usability problems (figure 1). In the figure the high problems are the most severe ones (these include e.g. if the user can not remove the application from the handset). Medium problems are not so severe, and low problems are not important ones. All the tests revealed problems, but the individual tests were most revealing.

CHI experts made the classification of problems and the criteria were that e.g. severe problems can prevent customers from using the service ("the costs of using the application") or can cause major problems to the user ("closing or removing the application").



	Total	High	Medium	Low
Total	25	3	9	13
Test I	24	3	9	12
Test II	16	2	6	8
Test III	11	2	3	6

Fig. 1. Test results / discovered problems

5 Conclusions

The pair-tests performed by non-professional moderators and carried out after focus group sessions seemed to reveal near half of the usability problems compared to individual usability tests, so several problems were reported *only* in single tests and not by user-pairs (8 problems). Most of these particular problems, which were reported only in single tests, were 'off-task' comments and these were not directly connected to the given test tasks (6 'off-task' problems). Surely off-task findings are also important input for the design work, but we think that in this kind of rapid design

the most important matter is to discover the basic task-related problems. Basically our findings suggest that with individual tests there appears to be more spontaneous feedback.

The most serious detail is that one of the severe problems ("closing the application") was not found in user-pairs tests. However, we suggest that our approach could be seen as an accurate enough method for the 'quick-and-dirty' testing in the busy mobile applications design field, especially if the iterative design principle is applied in the process (i.e. to test in more than 1 or 2 development phases).

Also, in individual tests there is probably some variation in results even with professional moderators [2,6], so we see that non-professional testers manage quite well when again we keep in mind the industry's limited needs.

An interesting point is that e.g. Hackman and Biers state that the user-pairs tests reveal more problems than individual tests [4] – our results do not support this.

Further studies will be needed in this area. We see this important, especially because there is really a great demand in the industry to develop 'light' design approaches or tools, which can be applied with time constraints and limited resources (financial, people). Involving non-professionals in the CHI activities may offer a solution in the pressing mobile application design work. After all, it is better to carry out imperfect usability tests than not to test at all.

References

1. Beyer, H. and Holtzblatt, K.: Contextual Design: Defining Customer-Centered Systems. Morgan Kaufmann Publishers, San Francisco, CA, USA (1998)
2. Boren, M.T. and Ramey, J.: Thinking Aloud: Reconciling Theory and Practice. IEEE Transactions on Professional Communication 43, 3 (2000) 261-278
3. Ericsson, K.A. and Simon, H.A.: Protocol Analysis: Verbal Reports as Data. MIT Press, Cambridge, MA, USA (1884)
4. Hackman, G.S. and Biers, D.W.: Team Usability Testing: Are Two Heads Better than One? Proc. of the Human Factors Society 36th Annual Meeting, HFES (1992) 1205-1209
5. Hom, J.: Co-discovery Method. http://jthom.best.vwh.net/usability/codiscvr.htm
6. Jacobsen, N.E. & John, B.E.: The Evaluator Effect in Usability Studies: Problem Detection and Severity Judgements. Proc. of the Human Factors and Ergonomics Society 42nd Annual Meeting, HFES (1998) 1336-1340
7. Wildman, D.: Getting the Most from Paired-User Testing. Interactions, 2,3 (1995) 21-27
8. Wilson, C. and Blostein, J.: Pros and Cons of Co-participation in Usability Studies. Usability Interface 4,4 (1998)

Pen-Based Gestures: An Approach to Reducing Screen Clutter in Mobile Computing

Mark Nicholson and Paul Vickers

Northumbria University, School of Informatics,
Pandon Building, Camden Street, Newcastle upon Tyne, NE2 1XE, UK
paul.vickers@northumbria.ac.uk
http://www.paulvickers.com/northumbria

Abstract. Mobile computing is an area of high growth despite having some serious design issues. It is difficult to increase the size of the screen because of the device's physical constraints. Consequently, as mobile applications have incorporated more functionality, screen clutter has increased. One method of reducing clutter is to remove visual controls and use pen-based gestures instead. We describe a cinema listing application for a Palm OS device that implements pen-based gestures as the main input method. Two methods are used to communicate the options available on each screen: audio cues and small visual prompts. Preliminary results suggest that buttons can be removed from the screen without detriment to task accuracy or user performance.

1 Introduction

As the computing power of mobile devices increases, so does the number of available applications. Already the organizational aspects of PDAs are being merged with the communication and entertainment aspects of mobile phones and portable music players. The move to one device with increased functionality presents a challenge to the user interface designer who is constrained by the physical limitations of mobile devices. Touch screens are the primary form of communication between the user and the PDA. One reason for this is the number of different applications found on PDAs. It would not be feasible to provide physical input widgets for all applications and keep the device small enough. The touch screen provides the developer with the ability to build virtual input widgets that occupy a percentage of the display area. Unfortunately, using the display in this manner deprives the interface developer of valuable output screen real estate. For example, following the Palm OS interface design guidelines [1] means a single button uses up to 10% of the available screen space (Fig. 1.)

To reduce clutter a number of different approaches can be implemented. Zooming interfaces have been incorporated into mobile computers (e.g. Halo [2], a zooming interface for a map viewer on an iPAQ). Sonification of widgets can reduce screen clutter. Adding earcons to the buttons of the Palm OS calculator application allowed a reduction in their size [4] from $5.0mm^2$ to $2.5mm^2$ with little real loss in usability (though an increase in subjective workload was observed).

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 320–324, 2004.

Palm OS uses the Graffiti® (and latterly the Jot®) gesture systems for data input. Application control is largely via buttons, though gestures are used for actions such as copy, paste, delete, etc. Mobile gestural input is usually pen-based, though spatial gestures have also been implemented (e.g. [5]). PDA users find gestures powerful, efficient, and convenient [7], but some find gestures difficult to remember and have problems with pen stroke recognition. In developing a gesture-based interface three factors must be considered. 1) gestures should be reliably recognized by the computer, 2) gestures should be easy for people to learn and remember [7], and 3) the user must be aware of the options that are available.

2 Pen-Based Gestures as an Input Modality

A mobile music player running on an iPAQ is a recent example of a pen-based gesture user interface [8] where simple, single-stroke gestures were used to control functionality such as next track, previous track and volume.To recognize a pen-based gesture a computer must track the pen/stylus, typically recording a set of coordinates as it moves over a flat sensing bed. Recognising a group of coordinates as a gesture can be done with a number of methods, including feature- and model-based techniques. A feature-based recognition system [9, 10] can be used on both static and dynamic character information. For example, a feature of the character “j” is the dot. The initial angle of the input stoke is a feature that could be used to differentiate between an “A” and “L” character. The path a “D” character takes through a dynamic bounding box, or grid, is a feature that could be used to differentiate it from a “B” character.

Model-based recognition algorithms are primarily implemented using hidden Markov models. An input stroke is segmented into a number of separate strokes. For example the character “A” could be segmented into two strokes “/” and “\” [6]. A powerful feature of stroke separation is the ability to predict the attempted character. The “A” character has two very distinct strokes. Therefore, following the input of the first stroke, the only remaining possible letters that could match the stoke are “A” and “X”.

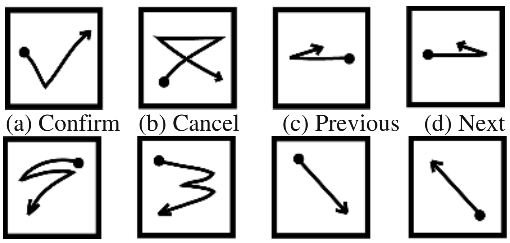
3 Gestures in a Cinema Listing Application

As a vehicle for demonstrating the use of sonic gestures a simple cinema listing application was built. The application allows the user to browse lists of movies showing at different cinemas and to make bookings and payments (Fig. 1.)

Six whole-screen gestures are required for the browsing and booking functions: ‘Next’, ‘Previous’, ‘Book’, ‘Showing’, ‘Confirm’ and ‘Cancel’ (see Fig. 2.) ‘Next’ and ‘Previous’ allow navigation through the application screen. ‘Showing’ lists the films that are currently showing. ‘Book’ places a booking for a movie. ‘Cancel’ and ‘Confirm’ have the usual meanings and are available when the user wishes to make a booking. ‘Backslash’ brings up a list of available gestures while ‘Forwardslash’ performed no function in the application but was used during user training.

(a) Normal screen (b) Buttons removed (c) Gesture prompts added

Fig. 1. Screen shots from the different versions of the application. Buttons take up to 10% of the available screen space

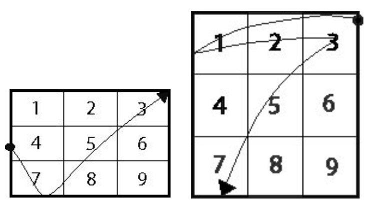


(a) Confirm (b) Cancel (c) Previous (d) Next

(e) Showing (f) Book (g) Forwardslash (h) Backslash

Fig. 2. Gesture groups: (a)-(d) Common functionality, (f)-(g) Cinema listing group

A simple feature-based recognition algorithm (similar to that used in [10]) was implemented to recognize the application's eight gestures. A dynamic bounding box divided into nine equal zones is drawn around the gesture (Fig. 3.) Each time a co-ordinate is captured, a counter for the co-ordinate's zone is incremented. To assist recognition of each gesture, the start and end zones of the gesture are also recorded. This information is matched against a library file and the associated function is executed. Problems are encountered with this approach when the input gesture has the same features as a stored gesture, but traverses different zones. To overcome this, tolerance was included in the library file allowing input gestures to drift into zones that are considered to be likely variations of the input gesture.



(a) "Confirm" (b) "Showing"

Fig. 3. Grid to fit two different gestures

The eight gestures are split into two groups. The common functionality group (Fig. 2a-d.) allows the user to perform navigation tasks and confirm or reject selections. The group is designed on a metaphor of the function performed by the gesture. The

Cinema Listing group (Fig. 2e-h.) allows the user to perform functionality tailored to the application. “Showing” and “Book” are primary gestures, allowing the user to view and book films showing at a selected cinema. “Backslash”, is a secondary gesture that causes a window showing the available input options to be displayed. “Forwardslash”, is secondary but not a part of the application–its purpose was to return the user to a start point during usability testing.

Two different methods were used to inform the user of the options available to them. The first, Sonically Enhanced Gestures (SEG), uses context-sensitive sound cues and the “Backslash” gesture. As the user navigates to a different screen, the cues associated with the available gestures are played in a serial fashion. The second, Sonically Enhanced Gestures with Permanent Screen Prompts (SEG-PSP), uses sound cues and small bitmaps, which are both context-sensitive (see Fig. 1c.)

Six audio cues were developed and the gestures divided into three groups. “Next” and “Previous” form a navigation group. “Confirm” and “Cancel” a decision group, and “Showing” and “Book” make up a cinema-specific group. Splitting the gestures into three groups allows a different number of tones to be the foundation of each group. Large spectral differences in pitch can then be applied within each group to aid the discrimination of each cue and help negate the audio limitations of the Palm V PDA on which the application was developed. For example, the navigation group cues use a single tone. “Next” is a D# (2488 Hz) and “Previous” a C (131 Hz). Gestures in the decision group use sequences of two tones, while the cinema specific group gestures have cues of three tones. To enable the user to distinguish where one cue finishes and another starts, a gap of 150ms is inserted between cues. The next stage of this research will involve a systematic design of structured earcons to represent the gestures–the focus of this stage was to implement working gestures.

The Cinema Listing application is a domain to test the effectiveness of gestures as a primary input widget. The application interface included a title, pick lists to navigate to a cinema, text fields to display cinema and film details and to accept booking and payment information, and widgets to view, book and make an electronic payment to watch a selection of films. Three versions of the application were built: Buttons mode, SEG mode, and SEG-PSP mode. Buttons mode uses only standard Palm widgets, such as pick lists and push buttons to navigate through the interface and perform functionality. The SEG and SEG-PSP modes replace the buttons with gestures and operate in the manner described above.

4 Discussion and Further Work

So far, only a brief evaluation with six participants has been carried out. Participants had to perform a number of browsing and booking tasks in the three different system modes. Initial results from the study are promising. For example, one subject found the Buttons mode easy to use but “…had to stop myself trying to use gestures, after a while buttons mode felt cumbersome”. Another tried to use the gestures when working in buttons mode suggesting that the gestures were easily learnt and readily accepted by users. Further development should see a fully sonically enhanced gesture implemented (one whose behaviour is represented aurally) not just the sonification of user options. Gestures remove the need to tap the stylus at a specific screen location. The benefits for the visually-impaired have not been explored here, but this could be a

worthwhile avenue of enquiry. Mobile computers could provide very useful functionality to this user group if the limitations of stylus-based input can be ameliorated. As functionality increases, so does the number of required gestures (think of the toolbar of a typical word processor). Participants had difficulty recalling the six earcons in SEG mode. Brewster [3] showed how earcon hierarchies could be used for navigating telephone-based interfaces in which there are several levels of nesting leading to many nodes of information. It would be beneficial to explore how a well-structured earcon hierarchy could be mapped onto an application command set.

References

1. http://www.palmsource.com/developers/. Last accessed March 2004.
2. Baudisch, P. & Rosenholtz, R. *Halo: a Technique for Visualizing Off-Screen Loactions.* In Proceeding of ACM Conference on Human Factors in Computer Systems (CHI'2003), Ft. Lauderdale, USA, ACM Press Addison-Wesley (2003), 481 - 488 .
3. Brewster, S.: Navigating Telephone-based Interfaces with Earcons. In: H. Thimbleby, B. O'Conaill, and P. Thomas, (eds.): People and Computers XII. Springer (1997) 39-56
4. Brewster, S. A. Overcoming the Lack of Screen Space on Mobile Computers. Personal and Ubiquitous Computing, (6)3, (2002). 188-205.
5. Harrison, B. L., Fishkin, K. P., Gujar, A., Mochan, C., Want, R.: Squeeze Me, Hold Me, Tilt Me! An Exploration of Manipulative User Interfaces. In: Proc. CHI '98 ACM Conference on Human Factors in Computer Systems (1998) 17-24
6. Li, X., Plamondon, R. & Parizeu, M. *Model-Based On-Line Handwritten Recognition.* In Proceedings of the 14th International Conference in Pattern Recognition, Brisbane, Australia, August (1998), 1134-1136.
7. Long, A.C., Landay, J.A., & Rowe, L.A. *Implications for a Gesture Design Tool.* Proceedings of ACM Conference on Human Factors in Computer Systems (CHI'99) (Pittsburgh, PA), ACM Press Addison-Wesley (1999), 40-47.
8. Pirhonen, A., Brewster, S.A. & Holguin, C. *Gestural and Audio Metaphors as a Means of Control for Mobile Devices.* In Proceeding of ACM Conference on Human Factors in Computer Systems (CHI'2002), Minneapolis, USA, ACM Press Addison-Wesley (2002), 291-298.
9. Rubine, D. *Specifying Gestures by Example*. Computer Graphics, (25)4 (1991) 329-337.
10. Worth, C. D. *xstroke: Full-screen Gesture Recognition for X.* Proceedings of the 2003 Usenix Annual Technical Conference (San Antonio, Texas), 9-14 June (2003) 187-196.

Dynamic Primitives for Gestural Interaction

Steven Strachan[1,2], Roderick Murray-Smith[1,3], Ian Oakley[2], and Jussi Ängeslevä[2]

[1] Hamilton Institute, NUI Maynooth, Maynooth, Ireland
steven.strachan@may.ie
[2] Media Lab Europe, Palpable Machines Group, Dublin, Ireland
{ian, jussi}@medialabeurope.org
[3] Dept. Computing Science, University Of Glasgow, Glasgow, Scotland
rod@dcs.gla.ac.uk

Abstract. We describe the implementation of an interaction technique which allows users to store and retrieve information and computational functionality on different parts of their body. We present a dynamic systems approach to gestural interaction using Dynamic Movement Primitives, which model a gesture as a second order dynamic system followed by a learned nonlinear transformation. We demonstrate that it is possible to learn models, even from single examples, which can simulate and classify the gestures needed for the *Body Space* project, running on a PocketPC with a 3-degree of freedom linear accelerometer.

1 Introduction

Mobile telephones, Personal Digital Assistants and handheld computers are currently one of the fastest growth areas of computing and this growth is extending into fully wearable systems. Existing devices have limited input and output capabilities, making them cumbersome and hard to use when mobile. Consequently, a current requirement in this field is the development of new interaction techniques specifically designed for mobile scenarios. One important aspect of interaction with a mobile or wearable device is that it has the potential to be continuous, with the user in constant, tightly coupled interaction with the system. In these scenarios, interaction need no longer consist of an exchange of discrete messages, but can form a rich and continuous dialogue.

The Body Mnemonics project[1] develops a new concept in interaction design. Essentially, it explores the idea of allowing users to store and retrieve information and computational functionality on different parts of their bodies. In this design, information can be stored and subsequently accessed by moving a handheld device to different locations around the body. This work addresses three problem areas in mobile computing: the high levels of attention required using the devices, the impersonal nature of their interfaces, and the socially exclusive modes of interaction they support.

The work described in this paper represents first steps to providing the technology to support the gestural interaction required by the body mnemonics

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 325–330, 2004.

concept. It is concerned with developing algorithms to infer the location of a handheld device. To provide a system that requires no additional equipment (such as worn tags or markers) to facilitate the identification of different locations, it relies on inertial sensing. Inertial sensing is a relatively new paradigm for interacting with mobile computers. Furthermore, it is a good example of continuous input; the device gathers information about user behaviour whenever it is being held or carried.

A number of researchers, such as Hinkley *et al.* [2] and Rekimoto [3], have demonstrated that inertial sensors can provide alternatives to the physical and on-screen buttons in handheld devices. They have described systems whereby shaking and tilting the device triggers different commands. However, these interfaces still possess a strong graphical component and little work has been conducted on 'screen-free' gestural interfaces. Pirhonen *et al.* [4] demonstrated a mobile mp3 player where gestures were sufficient to enable users to control the player without looking at the screen. We wish to develop the idea of screen free interaction to provide increased usability when 'on the move'.

2 Initial Explorations

Our initial investigations were conducted using an iPAQ5550 equipped with a 3-axis Xsens P^3C linear accelerometer attached to the serial port. We are concentrating on short trajectories originating and terminating at specific body locations. Several locations were considered as the source of each gesture. These were the left or right hip, where a device may naturally be held when not in use and the centre of the chest, where a device is often held to enable optimal viewing of its screen. To avoid issues of handedness, we chose to model our gestures as all originated from the centre of the chest.

Four body areas were chosen as gesture end points - left shoulder, right shoulder, back pocket and back of head. For the purposes of this exploration, all gestures were performed from the centre of the chest using the left hand whilst standing still.

The 'brute-force' approach of integrating the inertial measurements into positional trajectories and referring these to a spatial map of the body is not a strong option. A combination of uncertainty as to the precise initial position of the device and integration drift led to a substantial error margin. Figure 1 displays the trajectories inferred from acceleration measurements, for movements to the four different parts of the body, with 10 examples for each class of gesture, and makes clear the resulting inaccuracy at the end-points.

3 Dynamic Movement Primitives

The focus of this project was to choose a recognition algorithm that was flexible enough to model the required trajectories, but also constrained enough that it could be trained with minimal effort, using a small number of example gestures by a novice user.

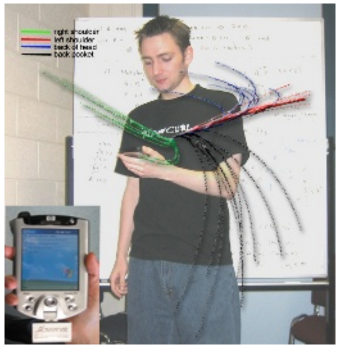


Fig. 1. Example of the drift encountered when acceleration traces are integrated into positions. Significant integration drift is observed, leading to end-point uncertainty.

The *Dynamic Movement Primitives* (DMP) algorithm proposed by Schaal *et al.*, is *"a formulation of movement primitives with autonomous non-linear differential equations whose time evolution creates smooth kinematic control policies"* [5,6]. The idea was developed for imitation-based learning in robotics, and is a natural candidate for application to gesture recognition in mobile devices. It allows us to model each gesture trajectory as the unfolding of a dynamic system, and is better able to account for the normal variability of such gestures. Importantly, the primitives approach models from origin to *goal* as opposed to the traditional point-to-point gestures used in other systems. This, along with the compact and very well-suited model structure enables us to train a system with very few examples, with a minimal amount of user training and also provides us with the opportunity to add richer feedback mechanisms to the interaction during the gesture.

DMP's are linearly parameterised enabling a natural application to supervised learning from demonstration. Gesture recognition is made possible by the temporal, scale and translational invariance of the differential equations with respect to the model parameters.

A Dynamic Movement Primitive consists of two sets of differential equations, namely a canonical system, $\tau\dot{x} = h(x)$ and a transformation system, $\tau\dot{y} = g(y, f(x))$. A point attractive system is instantiated by the second order dynamics

$$\tau\dot{z} = \alpha_z(\beta_z(g - y) - z), \;\; \tau\dot{y} = z + f, \tag{1}$$

where g is a known goal state (the left shoulder, for example), α_z and β_z are time constants, τ is a temporal scaling factor, y and $\dot{y}$ are the desired position and velocity of the movement and f is a linear function approximator. In the case of a non-linear discrete movement or gesture the linear function is converted to a non-linear deforming function

$$f(x, v, g) = \frac{\sum_{i=1}^{N} \psi_i w_i v}{\sum_{i=1}^{N} \psi_i}, \;\text{ where }\;\; \psi_i = e^{\left(-h_i\left(\frac{x}{g} - c_i\right)^2\right)} \tag{2}$$

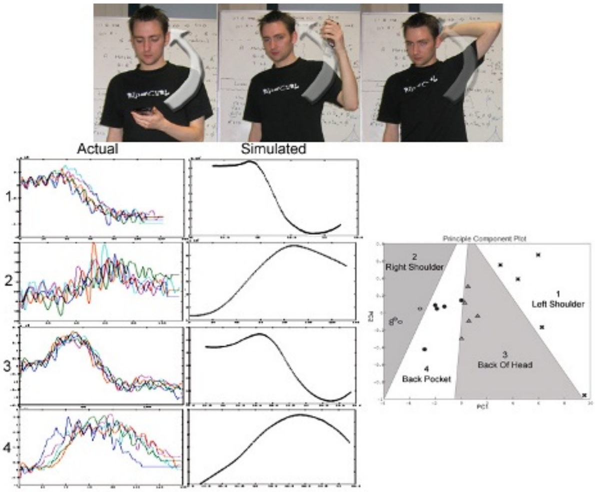


Fig. 2. Five realisations of the four gestures on the x-coordinate are shown along with an example simulated gesture from the DMP model. A principal component plot shows the separability of the model parameters. Similar results can be demonstrated for the y and z coordinates also.

These equations allow us to represent characteristic non-linear behaviour that defines the gesture, while maintaining the simplicity of the canonical 2nd order system driving it from start to goal. The transformation system for these discrete gestures is

$$\tau \dot{z} = \alpha_z(\beta_z(r - y) - z) + f, \;\; \tau \dot{y} = z, \;\; \tau \dot{r} = \alpha_g(g - r) \tag{3}$$

where $\dot{z}$, z and y represent the desired acceleration, velocity and position respectively.

The approach to learning and predicting the dynamic movement primitive is to provide a step change in reference and pass this through the non-linear deforming function. Values for the f's can be calculated along with sets of x's and v's from the canonical system and this is then passed through a Locally Weighted Projection Regression (LWPR) algorithm [7] that learns the attractor landscape and allows us to make predictions of the function f given values for x and v.

4 Results, Future Work, and Conclusions

Our implementation of the Schaal DMP algorithm, running on a pocket PC with inertial sensing, provides the basis for an efficient, robust and rapidly trainable gesture recognition system for the four basic gestures we tested.

Figure 2 shows examples of acceleration time-series corresponding to the x-coordinate acceleration trace for each class of gesture, along with the simulated

curve from the learned model in the second column. The good match between the measured and simulated curves provides encouraging evidence of its suitability for gesture recognition, especially as each simulated gesture was generated using a model trained on only one example of the five shown, and in only five iterations of the LWPR algorithm. The separability of the model parameters for classification purposes is visible in the plot of the first two principal components for each of the four classes of gesture.

The additional benefit of this dynamic approach is that it provides the designer with the opportunity to incorporate rich, continuous feedback mechanisms into the interaction with the user. We can now deliver continuous audio or tactile feedback relating to the user's motion, proximity to goals, or gesture trajectories [8]. We believe this kind of tightly coupled control loop will support a user's learning processes and convey a greater sense of being in control of the system. For this we will be using the MESH hardware platform [9], which features a 3-axis accelerometer, 3-axis gyroscope, 2-axis magnetometer and an integrated vibro-tactile transducer with a large (54 dB) dynamic range. The richer sensor input will broaden the scope of interaction possibilities, and the system features the dynamic vibrotactile output required to display the probabilistic feedback from our DMP models.

Acknowledgments. We acknowledge the HEA's support of the *Body Space* project, EC TMR grant HPRN-CT-1999-00107, EPSRC grant GR/R98105/01. SFI grants 00/PI.1/C067 and SC/2003/271.

References

1. Ängeslevä, J., Oakley, I., Hughes, S., O'Modhrain, S.: Body Mnemonics. In: Mobile HCI Conference 2003, Udine, Italy. (2003)
2. Hinckley, K., Sinclaire, M., Horvitz, E.: Sensing techniques for mobile interaction. In: ACM Symposium on User Interface Software and Technology, CHI Letters 2 (2. (2000) 91–100
3. Rekimoto, J.: Tilting operations for small screen interfaces. In: ACM Symposium on User Interface Software and Technology. (1996) 167–168
4. Pirhonen, A., Brewster, S., Holguin, C.: Gestural and Audio Metaphors as a Means of Control for Mobile Devices. In: Proceedings of ACM CHI2002 (Minneapolis, MN), ACM Press, Addison-Wesley (2002) 291–298
5. Schaal, S., Peters, J., Nakanishi, J., Ijspeert, A.: Learning Movement Primitives. In: International Symposium on Robotics Research (ISRR2003), Springer (2004)
6. Ijspeert, A., Nakanishi, J., Schaal, S.: Learning attractor landscapes for learning motor primitives. In S. Becker, S.T., Obermayer, K., eds.: Advances in Neural Information Processing Systems 15. MIT Press, Cambridge, MA (2003) 1523–1530
7. Vijaykumar, S., Schaal, S.: Locally Weighted Projection Regression: An $O(n)$ Algorithm for Incremental Real Time Learning in High Dimensional Space. In: 17th Inter. Conf. on Machine Learning (ICML2000), Stanford, California. (2004) 1079–1086

8. Williamson, J., Murray-Smith, R.: Granular synthesis for display of time-varying probability densities. In Hunt, A., Hermann, T., eds.: International Workshop on Interactive Sonification (Human Interaction with Auditory Displays). (2004)
9. Oakley, I., Ängeslevä, J., Hughes, S., O'Modhrain, S.: Tilt and Feel: Scrolling with Vibrotactile Display. In: Proceedings of Eurohaptics 2004, Munich Germany. (2004)

Touch Detection System for Mobile Terminals

Jani Mäntyjärvi[1], Kjell Nybergh[2], Johan Himberg[3], and Kari Hjelt[4]

[1] VTT Electronics, P.O. Box 1100, FI-90571 Oulu, Finland
Jani.Mantyjarvi@vtt.fi
[2] Neural Networks Research Centre, Helsinki Univ. of Technology, Finland
[3] Nokia Research Center, Helsinki, Finland
[4] Nokia Ventures Organization, Helsinki, Finland

Abstract. Knowledge when the terminal is in the hand of the user is important information that can be exploited in mobile applications. We present a touch detection system for mobile terminals based on impedance measurements. Experiments for recognizing touch of various objects are presented. The results show that the system is capable of recognizing if the device is touched with different objects such as bare hands, cotton (used e.g. in pockets of trousers) and leather carrying case.

1 Introduction

Augmenting the device with touch sensors makes it possible to develop input methods for triggering or initialising functions/applications in mobile terminals.

A study by Hinckley et al. presents touch sensing input devices, a mouse and a trackball, which use capacitive sensors to detect the touch [4]. In another study by Hinckley et al. they report sensing techniques for detecting when the device is in the user's hand or not [3]. Skin conductance (SC) measurement, often referred to as galvanic skin response (GSR), is a widely used method for measuring the electrical properties of human skin [6]. In SC measurement two electrodes placed near each other on the surface of the device measure the direct conductance of the skin.

Drawback with current touch detection systems is that they can give positive output when there is some other object e.g. metal or thin textile on the surface of the skin (e.g. pocket) in contact with the electrodes. This becomes a problem when touch detection is implemented on mobile device because they are placed e.g. pockets or bags.

Another and more accurate method for measuring electrical properties of the skin is skin impedance measurement, in which the frequency properties of the skin are examined with an AC-signal conducted into the skin using surface electrodes [7, 2].

We have applied skin impedance measurements for touch detection. Our motivation is that the method can be considered more reliable for detecting the presence of the hand, because it measures frequency dependent properties, which are characteristic of the skin. Also various objects can be detected because they have frequency dependent characteristics. We present an experimental implementation of a touch detection system for mobile terminals. The methods used are explained. Experimental

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 331–336, 2004.

results are provided. In the experiments we compare the performances of two types of basic discriminant analysis classifiers.

2 Touch Detection

Our approach is to first design suitable measurement system and secondly find expressive features for representing characteristic and recognise various classes by using the expressive features and a simple classifier.

Measurement of the electrical impedance of the skin. Skin impedance is considered to be high due to the high resistance of the outermost layer of the skin (stratum corneum) [7]. The stratum corneum consists mainly of dead skin cells. The thickness of stratum corneum is usually from 10 μm to 1 mm or more. The impedance of the skin is determined mainly by the stratum corneum at frequencies below 10 kHz [2]. The skin impedance is dependent on variables such as skin hydration, sensing pad size, and geometry. The potential and current difference between the two sensing pads is measured and the impedance between these two pads can be determined [7].

Feature extraction and pattern recognition. We evaluate the ability of the system to distinguish between different materials by applying simple classifiers to the data obtained from the measurements. Here, the concern is to find a compact set of expressive features. Moreover, the aim of the feature extraction and selection is to produce variables that discriminate the given classes. In this case, feature extraction methods are: 1) To find the center point of each curve (the average of points) and use the center point as a feature instead of all the points. 2) To fit a second-degree polynomial, $y=\mathbf{a}x^2+\mathbf{b}x+\mathbf{c}$, into the curve, and use its coefficients as features. We compare the results having either feature set 1 or 2. Basic classifiers, linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA), are used. Another rationale for using simple classifiers is that the data set is rather small. The risk of overfitting a powerful classifier and consequently making false conclusions is considerable.

Cross-validation (CV) is used in validating the classifier performance. Since we have only a few data samples, only few folds are used to guarantee that enough samples from every class are available in the training. A small number of folds is believed to give a pessimistic estimate and/or high variance for the classification power [5]. Therefore, we use a procedure referred to as Monte Carlo CV, where the CV procedure is repeatedly done several times [5].

3 Experiments and Results

Experimental setup. An experimental touch sensing system is implemented on a mobile terminal. We use two sensing elements, which are placed into the cover of the terminal (Fig 1). Experiments are performed with test persons holding the device in

their hand, with slightly moist cotton between skin and electrodes to simulate pockets e.g. of trousers. Measurements are also carried out when device is in a carry case in touch with leather. The sinusoidal output voltage with peak-to-peak range 600 mV and with frequencies 1, 2, 3, 4, 7, 10, 20, 30, 50, 70, and 100 kHz are used in a measurement. For each of the measurement, both the phase and amplitude of the output voltage U_o and U_i are determined and these are converted into imaginary and real parts. They are analysed using Matlab®. Offline analysis included classification experiments with LDA and QDA classifiers with CV. Classifiers are tested with different sets of features generated from the data.

The measurement setup is shown in Fig. 1 . It consists of a signal source, a resistor, a phase shift detector, and the mobile terminal with the two electrodes. The amplitude and phase of the input and output voltages are determined in phase shift detector.

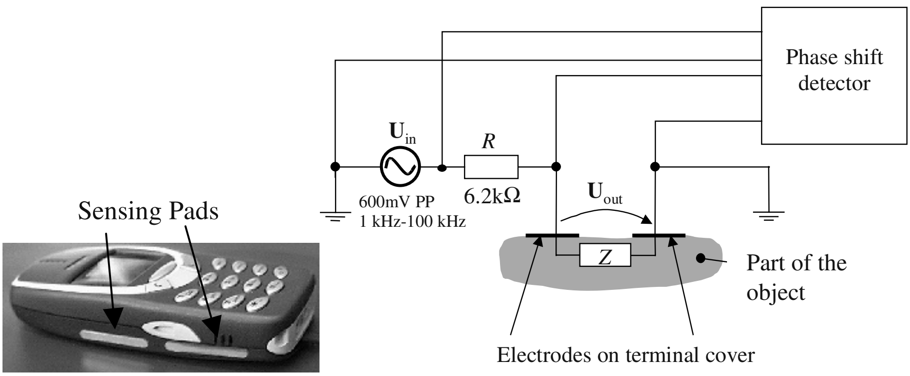


Fig. 1. Illustration of the implementation of sensing pads into a mobile terminal and schematics of the skin electrical impedance measurement setup

The measurements are performed at normal room temperature and humidity. Eleven male test persons were used. For each person the impedance is measured at the palm of the hand close to the thumb when the hand was bare. Additionally, measurements were carried out when there was a piece of 1) moist cotton between electrodes and skin, and 2) thick and moist leather from a phone carry case in contact with the electrodes. Two measurements with dry cotton and leather were carried out. In both cases it turned out that the resistive component was so high that no characteristic measurement values were obtained thus, it did not make sense to investigate these two alternatives any further. Moist materials are used to simulate humid conditions.

Measurement data. The total number of measurements (curves) is 33. Three measurements are presented as curves in Fig 2. Eleven points - each point obtained with certain frequency - from each measurement from a curve. Values of [$Im(U_o/U_i)$, $Re(U_o/U_i)$] measured from different objects differ from each other considerably. Values of moist cotton, illustrated by a solid line with circles, are considerably smaller compared to the others. The measurement values of skin, illustrated by a dotted line

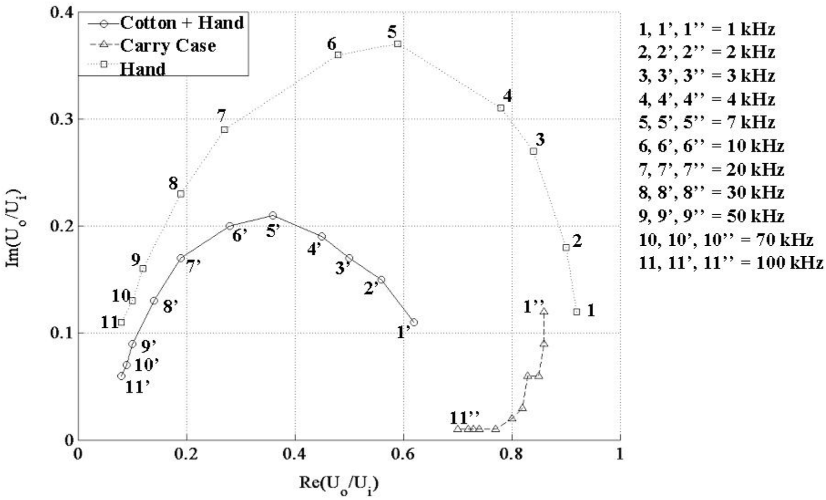


Fig. 2. The shapes of curves of hand and cotton are similar. The values of hand have larger dynamic range than values of moist cotton. The shape of curve from carry case measurements differs from the others

with squares, have larger dynamic range. The shapes of curves of skin and cotton are similar, whereas the shape of the curve obtained from leather, illustrated by a dashed line with triangles, differs from the others. Visual inspection of the measurements presented in Fig.2 shows that it is possible to find a set of features that leads to good classification results.

Classification Experiments. The classification experiments are performed in two phases. 1) The results obtained with various set of features are compared between the LDA and the QDA classifiers. 2) The results are examined with confusion matrix. A three-folded Monte Carlo CV is used in all classification experiments. The classifier is built three times and each time one group in turn is used only for testing while the two other groups are used for building the model. The classification with three-folded CV is performed 100 times for each classification experiment. The classification error and the values in confusion matrices are the averages of 100 classification experiments. The class averages and covariance matrices are maximum likelihood estimates.

Classification performances of the LDA and QDA classifiers are tested using a feature set consisting of center points of curves (averages of 11 points). The results with are presented in Tab. 1 on the left. The difference between the classification performance of QDA is negligible. The reason why the performance of QDA is not as good as LDA is that there is little data to estimate the covariance matrices, so the QDA will easily be overfitted. The classification performances of LDA and QDA are tested using the feature set consisting of polynomial coefficients (Tab. 1, on the right).

Table 1. Recognition accuracies obtained with various classifiers LDA and QDA.

	Feature set: Center points		Feature set: Polynomial coefficients	
	LDA	QDA	LDA	QDA
Cotton + Hand	100	96	72,4	99,2
Carrying Case	100	100	98,1	96,5
Hand	100	97	92,5	95,4

QDA provides better overall classification performance for this feature set. This might be because the class borders in this case are inherently nonlinear and QDA models them better. The recognition accuracy of LDA is good except for the recognition accuracy for the class *cotton + hand* 72.4%. In order to examine the performance of LDA classifications for features; polynomial coefficients, the confusion matrices are presented (Tab. 2). The confusion matrices are generated from averaging classification results obtained by using Monte Carlo CV. Tab. 2 shows that classes *cotton + hand* and *hand* mix systematically.

Table 2. Confusion matrix for classification results obtained with LDA classifier by using polynomial coefficients

Predicted class / Actual class	Cotton + Hand	Carrying Case	Hand
Cotton + hand	72.4	8.4	19.2
Carrying Case	1.9	98.1	0
Hand	7.42	0	92.5

The best results were achieved using center point features and LDA. In each test class *carrying case* and *hand* are classified with good accuracy. Data is measured in very stabile environmental and usage conditions with moist materials, which do not equal various unstabile usage situations of mobile terminals. When additional measurements are carried out in various usage and environmental conditions the collected data is variant and features can be examined more carefully. With a large amount of data the features that describe [Re,Im] curve more accurately might be relevant and improve the results.

4 Conclusions

An experimental implementation of a touch detection system for mobile terminals is presented. The system uses two sensor pads and impedance measurements for detecting the presence of an object. The discriminant classifier carries out the touch recognition. The performance of the system is examined by using two feature sets: central points and polynomial coefficients. The best classification results for recognition of objects: hand, cotton, and carrying case are almost 100% and they are obtained by using an LDA classifier with center point features. The comparison of the classifiers

is made tentatively without a rigorous analysis of statistical significance of the result. The results are good and show great promise of the chosen methods. In the future the performance of the system will be tested with larger data sets recorded in various usage situations.

References

1. Cherkassky, V., Mulier, F., Learning from Data, Wiley, (1998)
2. Grimnes S., Martinsen O.,G.: Bioimpedance and Bioelectricity Basics. Acad. Press (2000)
3. Hinckley, K., Pierce, J., Sinclair, M., Horwitz, E., Sensing Techniques for Mobile Interaction. CHI Letters 2(2), (2000) 91-100
4. Hinckley, K., Sinclair, M., Touch-Sensing Input Devices, In proceedings of Intl. Conference on Human Factors in Computing Systems, (1999) 223-230
5. Kohavi, R., A study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, Proc. of Int. Conference on Artificial Intelligence, (1995) 1137-1143
6. Lykken, D.T, Venables, P.H,: Direct Measurement of Skin Conductance: A Proposal for Standardization. Psychophysiology, Vol.8 No.5 , (1971) 656-672
7. Yamamoto, Y.: Measurement and Analysis of skin Electrical Impedance. In: Acta Derm Venereol Suppl. 185, (1994) 34-38

Mobile Text Input with Soft Keyboards: Optimization by Means of Visual Clues

Laurent Magnien, Jean Léon Bouraoui, and Nadine Vigouroux

IRIT, 118 Route de Narbonne,
31062 Toulouse, France
{lmagnien, bouraoui, vigourou}@irit.fr

Abstract. Soft keyboards are one of the most popular methods to input text for mobile pen-based computing. They allow text input to be performed through an onscreen graphical representation of a standard desk keyboard. Besides standard QWERTY keyboard layout, some researchers have proposed optimized alternative key organizations to improve user performances with soft keyboards. In this paper we propose and evaluate a solution using visual clues to facilitate the acceptance of these optimized layouts by novices.

1 Introduction

The emergence of mobile pen-based computing devices has largely affected text entry methods. Graphic tablets and Personal Digital Assistants, which do not incorporate a traditional keyboard, generally rely on alternative input systems like handwriting recognition or soft keyboard. Besides these two main technologies widely spread, many innovative methods such as [5], [6], [8], [10] (see [3] for an overview), are indicative of the activity of research in the text entry area since the development of mobile computing. This paper focuses on the study of particular soft keyboards: namely those with layout optimized to yield better performances than the standard QWERTY layout. Various soft keyboards like OPTI [1], METROPOLIS [9], or FITALY (www.fitaly.com) are based on this approach. Their designers usually rearrange letters in order to minimize pen travel during keyboarding. Compared to QWERTY keyboard, they have shown promising results for expert users. However recent studies [2], [7], point out that on the other side of the learning curve, the performance of novice users are limited by the lack of familiarity with the layout. Using an unfamiliar layout limits the helpfulness of any existing knowledge, because it implies a more systematic visual scan to find the keys.

In this paper, we propose to assist novice user in this process by means of visual clues. We think that the use of such clues when entering text can valuably save scanning time so that the users could be encouraged to adopt optimized soft keyboards. In the next sections we briefly present the solution that we propose to integrate to unfamiliar soft keyboards, and then we report in detail the experimental protocol we set up in order to evaluate this technique.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 337–341, 2004.

2 Use of Visual Clues

The solution that we propose relies, at each keystroke, on a double process: character prediction and highlighted display of the corresponding keys. In order to illustrate this, we implement a first prototype [4]. Our character prediction system is based on a French dictionary of 1462 words. It makes use of a lexical tree to determine, for each inputted character, the set of next possible characters.

To highlight the predicted character, we have opted for labeling the corresponding keys in bold, which is possible with a large range of keyboards. For example, figure 1 below illustrates a conceivable use of these visual clues with two different kinds of keyboard for the keyboarding of the word ‘interaction’. The user has already entered the prefix “inte” and the system proposes the next probable characters in bold type.

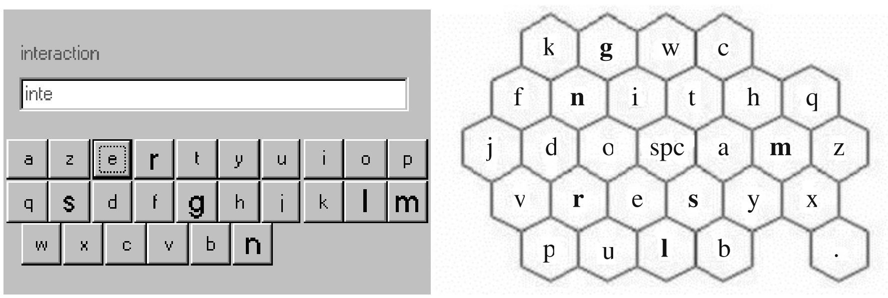


Fig. 1. Use of visual clues with an AZERTY (left) and a Metropolis-like keyboard [9] (right)

3 Experiment

We adopted an experimental approach to determine the influence of visual clues on the speed and accuracy of text input by beginners, with a soft keyboard. The main assumption was that visual clues under these conditions improve the text input rate. However we formulated a secondary assumption: this improvement tends to decrease when the prediction system is error-prone.

Twelve native French subjects were asked to enter three lists of words as fast as possible, by making as few mistakes as possible. They used a soft keyboard running under the following three modes:

- No visual clues (NVC mode);
- Visual clues with characters to be entered systematically among the highlighted ones (VC mode);
- Visual clues with, randomly, in 10% of cases, the character to be entered does not appear among the highlighted ones (VC10 mode).

Each list of 50 words was entered in a distinct mode. The last one (VC10) simulated a prediction system prone to error. Whatever the mode, we wanted to prevent the user from losing his novice status during the experiment. In this aim, we used the method presented in [2]. This method consists in randomly mixing the letter assignment to the keys after each inputted character. Thus the subjects were constantly faced with a new organization of keys. It should be noticed that this method defines the novice status by the absence of familiarity with a given organization of keys.

The experiment software was developed in C++ and is carried out on a Pocket PC. The software displays a keyboard according to the three different modes described above. The words to be entered appeared, one by one, in a textbox. They are selected randomly among a subset of French common nouns most frequently used in the child literature. This test set was selected because such words are presumed easy to spell. So that text entry errors due to misspelling would be avoided at the most. Words being entered appear in a second text box. The user's errors cause the emission of a "beep" sound, and the erroneous character is not entered.

The system of prediction integrated in the software provides on average 3.5 visual clues to the user.

4 Results and Discussion

The quantitative study of the collected data focuses on the text entry speed and the number of user mistakes according to the three experimental NVC, VC, VC10 conditions which refer to the exercises 1, 2 and 3 respectively. General results are shown in the table 1.

Table 1. Performances according to experimental conditions

	Exercise 1	Exercise 2	Exercise 3
Input time per character (s)	2.06	1.28	1.47
Error rate (%)	1.26	0.95	0.95

The error rate is low and does not vary much according to the experimental conditions. This simply indicates that the use of visual clues does not increase the number of user mistakes.

Effects of the Visual Clues on the Text Entry Performance

Results in the Table 1 show that the use of visual clues significantly saves time when entering a character. Figure 3 confirms this overall result for all the subjects tested. Time saved amounts to 37.7% when predictions are systematically right (VC mod). This improvement confirms our main hypothesis: use of visual clues does raise novice performance.

It would be helpful to assess this result in the light of other studies comparing the performances of an optimized soft keyboard with those of a standard one. [1] carried out an experiment over several sessions with the aim to compare the OPTI and the

QWERTY soft keyboard: During the first session the subjects typed 17 words per minute (wpm) on the OPTI soft keyboard while entering 28 wpm on the QWERTY soft keyboard.

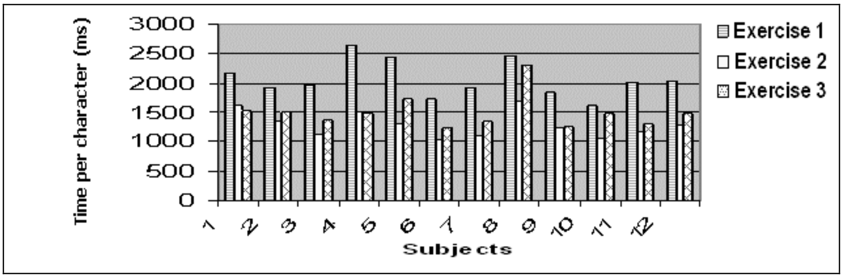


Fig. 2. Comparison of performances obtained by the 12 subjects with different exercises

Making the connection with the gain of time observed thanks to the use of visual clues is unfair because the protocols are different. However by applying to the gain of time observed during our experiment to the OPTI performances, the difference in the text entry speed between the two keyboards could be more than twice lower when resorting to visual clues. Even if this result does not derive from an empirical study, it is an incentive for the use of visual clues in the design of optimized soft keyboards.

Consequences of Prediction Errors

Notwithstanding erroneous visual clues, introduced in the VC10 situation of the exercise 3, the text input time measured is globally improved, i.e. 27.5% higher than in the NVC reference situation. It is 30% lower than in the PIV situation. In order to precise this result, let's further analyze the time spent under the VC10 mod, to enter a character when the letter is highlighted and when it is not.

Table 2. Performances in the VC10 situation

	With prediction errors	With correct predictions
Input time per character (s)	2.18	1.41

Table 2 shows that when the character to enter is not proposed among the highlighted candidate characters from the completion list, input time is about 6% higher than in the reference situation. Moreover when the right character is proposed, input time is about 11% higher compared to time in the VC situation, where the characters to input are systematically highlighted. These results confirm our second hypothesis at two levels: On the one hand because time input rises when the character to enter does not appear among the propositions; on the other hand because the system errors hamper the efficiency of visual clues for the remaining text entry. This proves the crucial role of the prediction system performances but does not cast definitive doubt over the use of visual clues as far as, in spite of 10% of erroneous propositions, they do significantly improve text entry speed.

5 Conclusions and Perspectives

This paper aimed to present a solution for optimizing unfamiliar keyboard layouts. It consists in using a prediction system to put into contrast the keys most likely to be typed. The keys are highlighted in a way that ensures layout-independence. Our primary assumption suggested that this method would grow up novice text entry rate. That could lead, partly at least, to better acceptance of soft keyboard layouts which are optimized but also unfamiliar.

We carried out an experiment, which confirms this assumption: correct predictions nearly lead to a 40% gain in speed. Another important result underlines the effects of prediction system performances. Error-prone system deteriorates user performance but does not necessarily destroy the positive effect of the recourse to visual clues: in spite of 10% of errors of setting in contrast, they provide one still significant improvement.

However promising these results are, they remain preliminary to a more in-depth study, which shall make it possible to specify the influence of the number of visual clues and different error rates of prediction on user performances. Furthermore, this study shall analyze the effect of visual clues on the transition from novice to expert status. The expected results shall allow to provide recommendations for the use of visual clues in the design of optimized layout for soft keyboards.

References

1. MacKenzie, I. S., & Zhang, S. Z., The design and evaluation of a high performance soft keyboard. Proceedings of CHI '99, New York: ACM, 1999, pp.25-31.
2. MacKenzie, I. S., and Zhang, S. X., An empirical investigation of the novice experience with soft keyboards. Behaviour & Information Technology, 20, 2001, pp. 411-418.
3. MacKenzie, I. S., and Soukoreff, R. W., Text entry for mobile computing: Models and methods, theory and practice. Human-computer interaction, 17 (2002), 147-198.
4. Magnien, L, Bouraoui, J.L., Vella, F., Uilisation d'indices visuels pour l'aide à la saisie de texte sur PDA, IHM'03, 2003, pp. 252-255.
5. Mankoff, J., Abowd G. D., Cirrin: A word-level unistroke keyboard for pen input, Proceedings of UIST'98, ACM, pp. 213-214.
6. Partridge, K., Chatterjee, S., Sazawal, V., Borriello, G, Roy, W., TiltType: Accelerometer-Supported Text Entry for Very Small Devices, Proceedings of UIST'02, pp. 201-204, November 2002
7. Sears A., Jacko J. A., Chu J. and Moro F., The role of visual search in the design of effective soft keyboards, Behaviour & Information Technology, 2001, 20, pp. 159-166.
8. Ward, D., Blackwell, A., MacKay, D., Dasher – A data entry interface using continuous gesture and language models. Proc. ACM UIST. 2000. p. 129-136.
9. Zhai, S., Hunter, M. and Smith, B. A., 2000, The Metropolis keyboard: An exploration of quantitative techniques for graphical keyboard design, Proceedings of UIST'00, (New York: ACM), pp. 119-128.
10. Zhai, S., Kristensson, P-O., Shorthand Writing on Stylus Keyboard, in Proceedings of CHI 2003, Fort Lauderdale, Florida, April 5-10, 2003.

Watch-Top Text-Entry: Can Phone-Style Predictive Text-Entry Work with Only 5 Buttons?

Mark D. Dunlop

Computer and Information Sciences, University of Strathclyde, Glasgow G1 1XH, Scotland
Mark.Dunlop@cis.strath.ac.uk

Abstract This paper presents an initial study into the viability of text entry on a watch face using four alphabetic buttons and a central space key. The study includes a technical evaluation of likely error rates using a large text corpus and user studies on palmtop emulated mobile phone and watch. The results, though in favour of the phone pad, are encouraging and show such a method is feasible.

1 Introduction

Predictive text-entry on mobile phones, as standardised by Tegic's T9 software [1], has proven extremely effective for mobile phone keypads [e.g. 2, 3]. However, this method still requires a keypad of 9 buttons (8 alphabetic and 1 space for plain text entry). In this paper we report our initial investigation into using a 5 key pad for predictive text entry targeted at watch-top text-entry. The pad used here consists of four soft alphabetic keys around the periphery of a touch screen and a central space key (see fig 1). The motivation is to allow relatively high speed text entry on very small device using an approach familiar to mobile phone users (c.f. very small keypad designs such as [4]) and without the need for a stylus (c.f. handwriting (e.g. Graffiti), many-key soft-keyboards (e.g see [3]), or gesture input (e.g. T-Cube or Cirrin [6]).

Predictive text-entry is based around a large dictionary of word senses with occurrence information, users press one key per letter from multiple-letter keys and the system suggests possible matches to the key sequence in descending occurrence frequency. The simplified text entry approach used here overloads the space key: on first press a space is entered, on subsequent consecutive presses the suggested word cycles. For example to enter *LUNCH* using the interface in figure 1, the user would press *NMLKJIH*, followed by *UVWXYZ*, *NOPQRST*, *ABCDEFG* then *GHIJKLM* at which point *HUMAN* would be suggested as the most common word from those five keys, the user would press space to enter a space followed by another space to cycle words resulting in *LUNCH*. Predictive text-entry methods inherently have a level of errors – there are often more than one word possible from a given key sequence. While presenting words in

Fig. 1. 5-key text entry

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 342–346, 2004.

decreasing order of occurrence frequency reduces the commonality of errors, they still occur. When reducing from eight to four alphabetic keys it is expected that the number of errors will increase. To assess how much the error rate increases a technical experiment was conducted and is report here. Having fewer keys also implies users have fewer, larger, targets to hit and, in fig1, these are centred around *space* making a very close set of relatively large targets. Following Fitt's law, we may expect faster interaction these buttons. To assess use of the keypad, user experiments were run measuring input speed and error rate and are reported later.

2 Technical Experimental Setup

The technical experiments were based around a dictionary of 77 317 word senses, with frequency information, extracted from six months of *The Herald* newspaper (same as in [2]). The performance of encoding an individual word is dependent both on the keypad layout and the dictionary and was measured as follows:

$$p_{w,d,k} = \frac{|k(w,d)|}{|w|}$$

where $|k(w,d)|$ is the length of the encoding word w by keypad k using dictionary d.

The performance for the top n words was calculated using a weighted average, by frequency of occurrence, of each word in the top n, as follows:

$$P_{n,d,k} = \frac{\sum_n p_{w_n,d,k} . f(w_n, d)}{\sum_n f(w_n, d)}$$

where $f(w_n,d)$ is the frequency of occurrence of word n in dictionary d.

Using *The Herald* dictionary P_{200} was calculated for the six possible balanced alphabetic keypad layouts using four buttons, to assess the best keypad layout for alphabetic ordering. This analysis resulted in the keypad: *ABCDEFG*, *HIJKLMN*, *OPQRST* and *UVWXYZ* being used as the alphabetic ordered keypad.

Of course, letters do not need to be distributed alphabetically and a separate study was conducted to estimate the best possible key layout from the 4^{26} possible keypads. All 2, 3, and 4 letter words in the dictionary were evaluated to assess the pairwise confusion of individual letters based on one letter error per word, i.e. a measure of how likely swapping one letter for an other would result in a valid word. This resulted in a table[1] of 325 confusion weights, which were sorted into decreasing confusion occurrence to give *AI*, *ST*, *NS*, *NT* and *IO* at the top. Each of the four alphabetic keys was initially assigned one letter from *AIST* and their running total of confusion weights set to zero. For each subsequent letter from the list of pairs that had not already been assigned, a potential confusion weight was calculated as the sum of all confusion weights for combinations of letters currently on the key plus the new letter. The new letter was then added to the key with the smallest resulting total confusion score to minimise the total confusion weight per key (e.g. *N* is added to the *I* key as the confusion weight between *NA*, *NI*, *NS*, and *NT* is lowest for *NI*). This process resulted in the *GORSUV* keypad with the following four keys (rearranged

[1] See http://www.cis.strath.ac.uk/~mdd/research/files/confusionscores.html

alphabetically): *GORSUV*, *AFKMWXY*, *BDILNQZ* and *CEHJPT* (see figure 2). The *GORSUV* keypad was then used as an estimated optimal keypad.

Finally, for comparison a similar scheme was used for the traditional mobile phone keypad using both predictive text entry and multi-click text-entry (using the multi-click encoding instead of *k(w,d)* but weighting similarly to the dictionary methods).

Fig. 2. GORSUV key-pad

Table 1. Weighted keys per letter for different keypads and number of top 50/200 words that appeared as first choice on list of suggested words when entered

keypad	P_{200}	*top 50 as 1st hit*	*top 200 as 1st hit*
multi-click phone	2.101	n/a	n/a
alphabetic watch	1.060	45	162
GORSUV watch	1.041	46	166
predictive phone	1.009	50	191

Table 1 shows that, on a weighted average over the top 200 words in *The Herald*, the predictive phone keypad achieves an impressive average of 1.009 keys per letter. The *GORSUV* and alphabetic keypads perform significantly worse than the phone pad with 1.041 and 1.060 keystrokes per letter, while multi-click entry averages to over twice as many keystrokes per letter. Table 1 also shows how many of the top-50 and top-200 most common words were suggested as first match when keyed in.

While performing worse than a mobile phone, the suggested error rates for both *GORSUV* and alphabetic four-key pads are encouragingly good and not as bad as may be expected from halving the number of alphabetic keys. While on both measures, *GORSUV* is better than alphabetic it is not clear whether the much longer training time for *GORSUV* would be worth the effort.

3 Usability Experimental Setup

Usability experiments were conducted on a touch sensitive iPAQ handheld computer with phone (fig 3) and watch (fig 1) simulations written in Java using the same dictionary. Due to memory limitations of handheld Java the dictionary was limited to the top 9000 words from the 77k dictionary used above (augmented with 6 out-of-dictionary words).

Fig. 3. Phone Emulation

The experiment followed a within-subject design with two training and two timed task-sets per subject. Each of the four task-sets was composed of entering 3 sentences, from an independent list of humorous short phrases[2], on one interface. The experiment was balanced for first-use system and first-use task-set. Subjects were timed and errors recorded. Twelve subjects carried out the test in total, mostly MSc and PhD students in Computer Science plus

[2] http://www.pbbt.com/Directory/Jokes/681.html

two lecturers. The interfaces deliberately did not include a backspace, to remove correction time from timings, instead users were instructed to hit space and move on to the next word.

Table 2 shows the times for the whole timed task sets averaged over all users, together with the times for just the last two sentences (timing varied more over the first sentence as the user settled with the device). Table 2 also shows the number of words incorrectly entered for each device.

Table 2. Timing and total error count results from user trials (significant results in bold)

	Watch		Phone	
	mean	stdev	mean	stdev
3 sentences	**3.87**	0.89	**2.75**	0.59
2 sentences	**2.18**	0.60	**1.41**	0.47
Errors	0.75	0.87	1.17	1.08

Not surprisingly, the results show statistically significant faster performance with the phone keypad over the watch for both 3 and 2 sentence statistics (at 1% one-tailed correlated t-test). The table also shows no significant difference in error rate between Phone and Watch interfaces. Errors were generally very low, with most errors being caused by a misspelling of a word resulting in wrong suggestions. When asked all users stated that the interface response was suitably fast and did not hinder their interaction.

Over the three sentences the watch was on average 40% slower. Given that many subjects commented that they would expect to get better over time as they still felt they were learning the keypad, this is not a surprising result and shows that the watch keypad, while not reaching the performance of a phone keypad, would be usable for text entry. All subjects stated that they would use the phone in preference to the watch, but that (in all bar two cases, where the subject did not wear a watch) they would sometimes use the watch if given one. One subject highlighted that if holding the watch, two-thumbed text entry could be extremely fast and comfortable.

4 Discussion

The study reported here was on a short timescale (around 30 mins per subject), a longer trial would be needed to fully assess the speed of entry as it is clear users had not reached a comfort level with the watch interface (and many were very fast phone texters). Ideally the system for subsequent trials would be implemented on a real touch-sensitive watch to assess long-term usage. The use of a newspaper also biased the language somewhat differently to that of normal text messaging, e.g. *lunch* is likely to be more popular than *human* in text messaging. However, the dictionary was used comparatively throughout so this is unlikely to affect results here but would need to be replaced for a long-term study.

The current implementation of watch-face text-entry does not support capitalisation, punctuation, error correction or menu commands. These would have to be implemented using a combination of gestures, two-finger chords, long presses or physical buttons on the side of the watch. Investigations are planned to develop and test a full text entry method for small screens based around the interface presented here. The use of overloading space, almost required for the watch interface, did not

cause any usability problems even for very regular texters. However, this might not be the case when complex schemes are needed to replace the automatic space with punctuation marks etc.; again further investigation is required. The current watch interface does not have “dead-zones” between keys, which may be explain some common misspellings (e.g. users attempting to enter *g* and hitting the *H-N* key instead); while dead-zones would reduce the target zone size it may increase accuracy and requires investigation.

One final improvement that will be investigated is a variant of key-blanking techniques often used on scanning keyboards for people with severe motor control difficulties. These soft-keyboards often omit letters that do not occur in next position of a sequence. In the watch interface greying out the letters on the watch face that are not valid would not change the functionality or timing directly, as there are a fixed number of keys, however it might help users search time for the right letter.

The results presented here show that the use of a five-key keypad does increase the number of times a user needs to scroll down a list of suggested words for predictive text entry. However, this increase is not as great as may be expected by reducing the alphabet to only four keys. The paper presented the *GORSUV* keypad, a pseudo-optimal key arrangement of four keys. While this keypad does have better performance, results here show this improvement to be small and thus unlikely to be of benefit to all bar very frequent users given the extra time needed for users to find the correct key. User trials confirmed that the watch keypad was slower than the phone keypad, though again by not as much as might be expected (approx 40% slower than a touch screen emulation of a phone, however, this itself likely to be slower than using a physical keypad on a real phone). Furthermore, many users stated that they would expect to improve with regular use.

Overall, the results are encouraging and while the watch interface is confirmed to be slower than a phone interface for text entry, the results show that text entry speed on a watch-face by a frequent user can be expected to be reasonably close to that on mobile phone keypad. Furthermore, users were all comfortable with the text entry method after very little training, satisfying the need for a method similar to mobile phones.

Acknowledgements. My gratitude is extended to the subjects for giving their time to these experiments.

References

1. Kushler, C., “AAC Using a Reduced Keyboard”, *Proceedings of Technology and Persons with Disabilities Conference*, California, 1998
2. Dunlop, M. D., and Crossan, A., "Dictionary based text entry method for mobile phones", *Proceedings of 2nd Workshop on HCI with Mobile Devices*, August 1999.
3. Silfverberg, M., MacKenzie, I. S., & Korhonen, P. “Predicting text entry speeds on mobile phones”. *Proceedings of the ACM CHI 2000*, 2000.
4. MacKenzie, I. S. “Mobile text entry using three keys”. *Proc. NordiCHI 2002*, 2002.
5. Venolia, D., Neiberg, F., "T-Cube: a fast, self-disclosing pen-based alphabet", Proc. ACH CHI 94, 1994
6. J. Mankoff and G. Abowd. Cirrin: A World-Level Unistroke Keyboard for Pen Input. In *Proceedings of UIST'98*, 1998.

Pair-wise Variability Index: Evaluating the Cognitive Difficulty of Using Mobile Text Entry Systems

Frode Eika Sandnes[1] and Hua-Li Jian[2]

[1] Oslo University College, P.O. Box 4, St. Olav's plass, N-0130 Oslo, Norway
[2] National Cheng Kung University, Daxue road 1, Tainan City 70101, Taiwan

Abstract. A modified pair-wise variability index for evaluating the cognitive difficulty of using mobile text entry systems is proposed. The index is easy to compute from keystroke logs acquired from typing experiments where keystroke times are recorded. The effectiveness of the pair-wise variability index is demonstrated on the keystroke logs acquired using three different text entry strategies.

1 Introduction

In recent years text entry research has received renewed interest with the emergence of the mobile computing paradigm. Users take their information systems everywhere by the means of miniature mobile devices. When using certain applications, for example messaging, the users are required to enter text on their personal devices. However, small devices have no room for full sized QWERTY keyboards and much of the current research focuses on novel text entry strategies for effectively entering text on small resource-limited devices. Some research also addresses the evaluation of such systems. Most of the evaluation metrics are performance-oriented, i.e. number of characters per second or words per minute — ultimately, fast text-entry is the objective. Some studies also address errors and various error metrics [1]. Another line of research considers keystrokes per character (KSPC) [2] — a quantity computed theoretically without tests. However, several studies point out the inaccuracies and inadequacies of KSPC as it does not reflect actual performance [3]. In early work on chord keyboards Gopher and Raij [4] pointed out the importance of modeling the cognitive aspects of the text entry process. In this paper we are introducing the pairwise variability index, which is an attempt at revealing cognitive factors affecting text entry performance. The pairwise variability index was first introduced by Low, Grabe and Nolan [5] for comparing rhythm of speech through acoustic vowel measurements and has later been widely applied to other languages in the field of acoustical phonetics.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 347–350, 2004.

2 Cognitive Aspects of Mobile Text Entry

Most research into mobile text entry focuses on either increasing text entry speeds by reducing the physical time to enter text, by keyboard layout optimizations [6] or in some other ways minimizing the number of operations or keystrokes needed to retrieve characters (KSPC minimization [3]). However, few of these models address cognitive aspects of the typing task. In this paper we make the assumption that the cognitive processing delay is somehow related to the inter-keystroke delay, i.e. the time duration between two consecutive keystrokes. After a user has completed typing a character the user wishes to type the next character. In order to enter the character the user needs to perform one or more operations depending on the particular text entry strategy. If the delay between the keystrokes is short the task is more likely to be easy than when this delay is long. Based on inter-keystroke delay measurements a centrality measure can be computed as either a mean or a median. Some researcher prefer the median as this measure of centrality is more robust to outliers in the data set, and outliers are quite common in keystroke logs. Users occasionally take short breaks, they need time to decide or read what to type next or their thoughts simply wander off. Therefore, typing experiments often comprise passages of quite regular, or close to regular, rhythmical patterns of keystroke events and scattered intervals of irregular delays, both in terms of start time and duration. However, simply using the mean inter-keystroke delay one may not capture the necessary information. We hypothesise (but do not attempt to prove) that usable text entry techniques allow users to type in a rhythmical manner, while a poor text interface will results in irregular keystroke rhythms.

3 Inter-keystroke Delays and the Pair-wise Variability Index

Given a log-file obtained from a typing experiment comprising a set of keystroke times $t_1, t_2, .., t_n$ in some time-unit, where n is the number of keystroke measurements. The inter- keystroke delay $d_i, i \in [1..n-1]$ between keystroke timestamps t_i and t_{i+1} is therefore simply:

$$d_i = t_{i+1} - t_i \tag{1}$$

The normalized pairwise variability index $npvi_i$ of two consecutive inter-keystroke delays d_i and d_{i+1} for $i \in [1..n-2]$ is computed as

$$npvi_i = \frac{|d_i - d_{i+1}|}{d_i + d_{i+1}} \tag{2}$$

In this paper we refer to the normalized pairwise variability index as simply the pairwise variability Index. The original pairwise variability index proposed in [5] was expressed as the mean of the pairwise variability indices. However, since logfile datasets often contain some severe outliers that can significantly

Table 1. Results of the typing experiment. ikd = inter keystroke delay, npvi = normalised pair-wise variability index

Subject	Measure	MultiTap	Tree-based	One-stroke
Subject # 1	median ikd	0.5	1.4	1.2
	mean chars/min	22.5	13.0	28.5
	median npvi	0.29	0.55	0.41
	preference	3/5	0/5	4/5
Subject # 2	median ikd	0.52	1.02	1.94
	mean chars/min	27.2	18.6	31.1
	median npvi	0.25	0.46	0.32
	preference	1/5	5/5	4/5
Subject # 3	median ikd	0.24	2.13	0.55
	mean chars/min	26.5	7.7	26.2
	median npvi	0.39	0.49	0.46
	preference	4/5	0/5	3/5

affect the overall mean we propose instead to represent the pairwise variability index in terms of the median of all the individual pairwise variability indices, namely the median of the values $npvi_1$, $npvi_2$, ..., $npvi_{n-2}$.

We hypothesise that a text entry strategy which yields a low pairwise variability index indicates that it is easier to use than a text entry strategy with a higher pairwise variability index.

4 Experimental Evaluation

Our assessment is based on the data from an experiment reported in [7] where three subjects were asked to type text using three one-handed five-key text entry strategies. The subjects were asked to practice for five minutes and the keystrokes for the following 15-minute typing session were recorded. The three techniques consisted of a five-key multi-tap technique analogous to the multi-tap technique found on most mobile handsets, a five key tree-based method where letters are retrieved in two steps from a two-level hierarchical menu system and a one-stroke approach similar to the T9 text entry, but with only five keys. Statistics based on this experiment are listed in Table 1.

Table 1 shows the median inter keystroke delay (ikd), i.e. the time between two consecutive keystrokes, mean number of characters typed per minute, median normalised pairwise variability index and users' indication of preference based on a questionnaire.

Clearly, the multi-tap method has the smallest pairwise variability index, followed by the one-stroke method and finally the tree-based method. The small variability for the multi-tap method is consistent with the fact that it is the easiest method to learn and use. The user simply scrolls through the characters by tapping the keys, and often the same key is pressed repeatedly in sequence. The one-stroke method results in a larger variability than the multi-tap method as the user needs time to decide which of the five keys to press in order to retrieve

the next character and the tree-based method results in the largest variability since the user must make a decision on which of the five keys to press twice for each characters.

Two of the subjects prefer the multi-tap and one-stroke method to the tree-based method, while one subject preferred the tree-based and the multi-stroke method to the multi-tap method. The preference in all cases for the one-stroke method can be explained from the fact that it is the most productive method resulting in the highest number of characters per minute. Subject 2 who preferred the tree-based method over the multi-tap method reported that he lost patience with the repeated tapping in the multi-tap method and this may have affected his opinion.

Note that for the one-stroke method there are also delays associated with selecting words from lists when there are ambiguities. These delays are not taken into consideration as they are filtered by the median measure of centrality.

5 Conclusions

The pair-wise variability index of inter-keystroke delays extracted from keystroke logs were proposed as an experimental quantity for comparing the effectiveness of mobile text entry strategies. Our preliminary investigation of the pair-wise variability index applied to a typing measurement shows that the pair-wise index is a useful indicator of how users respond to a text entry strategy. However, the index should not be used to rank text entry strategies exclusively, but it should rather be used as a complimentary measure in conjunction with other observations and statistics. The pair-wise variability index is easy to compute and is robust to outliers. Further, it is independent of the time units used in the measurements and the relative typing rates of individual subjects.

References

1. Soukoreff, R.W., MacKenzie, I.S.: Metrics for text entry research: An evaluation of msd and kspc, and a new unified error metric. In: Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI 2003, pp. 113-120. New York: ACM. (2003)
2. MacKenzie, I.S.: Kspc (keystrokes per character) as a characteristic of text entry techniques. In: Proceedings of HCI Mobile 2002. (2002)
3. MacKenzie, I.S.: Mobile text entry using three keys. In: ACM proceedings of NordCHI 02. (2002) pp27–34
4. Gopher, E., Raij, D.: Typing with a two-hand chord keyboard: Will the qwerty become obsolete? IEEE Trans. on Systems, Man and Cybernetics **18** (1985) 601–609
5. Low, L.E., Grabe, E., Nolan, F.: Quantitative characterizations of speech rhythm: Syllable-timing in singapore english. Language and Speech **43** (2000) 377–407
6. MacKenzie, I.S., X., Z.S.: The design and evaluation of a high-performance soft keyboard. In: ACM proceedings of CHI99. (1999) 25–31
7. Sandnes, F.E.: One-handed mobile text-entry: evaluation of five-key text entry techniques. In: Proceedings of MOBIS'04, Kluwer, Oslo, Norway (to appear). (2004)

Xaudio: Results from a Field Trial Study on a Technology Enhancing Radio Listeners' User Experience

Reinhard Sefelin, Verena Seibert-Giller, and Manfred Tscheligi

CURE Center for Usability Research & Engineering,
Hauffgasse 3-5,
1110 Vienna, Austria
sefelin@cure.at

Abstract. This paper discusses a field trial of a technology (Xaudio), which connects two different types of media: The mobile internet and the radio. Inaudible signals (watermarks) are broadcasted via the sound of the radio, are received by a mobile device and decoded. This information then is used to take the listener directly to a mobile application (Xaudio application) that is relevant to the radio content currently broadcasted. Ten persons participated in a field trial study on this technology. The paper presents the results of this field trial. In particular it compares different kinds of applications and analyses the reasons for their success or failure. Furthermore proposals to improve the service of Xaudio for future uses are discussed briefly.

1 What Is Xaudio?

Xaudio is an active service entailing the insertion of inaudible codes (watermarks) into the broadcasted audio. These codes survive broadcast, transmission through the air between speakers and microphones, and can be extracted in real-time by portable mobile devices such as mobile phones and Personal Digital Assistants (PDAs). The extracted codes, which uniquely identify both the broadcaster and the content being played, can then be used to enable the listener to access so called Xaudio applications. These applications provide content that is related to the radio programme that is currently broadcasted. For more information on Xaudio see [3].

2 Trial Set Up

2.1 Decoder Interface and Xaudio Applications

Over a period of 13 days Radio Hit (a Slovenian radio station broadcasting in the area of Ljubiljana) broadcasted watermarked programme. During two hours in the morning and in the afternoon, the whole content of the radio programme was "watermarked" and associated with Xaudio applications. The Xaudio applications "News",

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 351–355, 2004.

"Traffic", "Weather" and (to a certain degree) "Commercials" were available and updated during the full period (24 hours a day). The number of Xaudio applications associated with songs was limited to 200. Songs, for which no separate Xaudio application was available, led to the start page of Xaudio (see Figure 1).

For the whole trial period each trial participant was equipped with a Xda II [2]. The Xda II is a pen-based PDA including an integrated personal digital organiser and a mobile phone. The ten participants (average age: 26 years) were informed that they can use the device for all different kinds of purposes (making telephone calls, surfing the internet) and that they should not worry about costs.

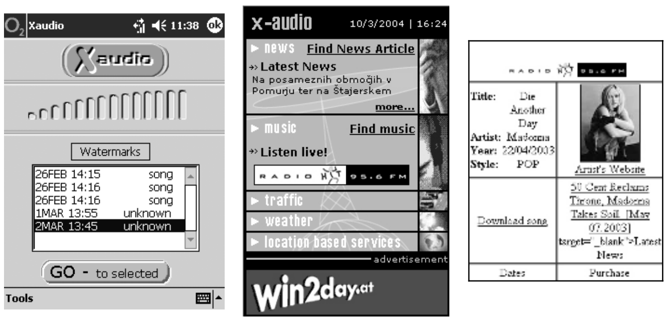


Fig. 1. Left: Decoder Interface (Main window); Middle: Xaudio Application (Start page); Right: Xaudio Application (Music)

On each device Xaudio's decoder interface was implemented. The interface enabled users to see new watermarks arriving and to open the corresponding applications. Apart from these main functions the decoder interface enabled users to filter watermarks and to choose whether a new watermark should be followed automatically or only after hitting the "Go" button. Figure 1 shows the main window of the decoder interface, the start page of the Xaudio applications (all applications could also be reached from this page) and an example of a music application. One of the commercial-applications included a location-based service (LBS). The watermark that was associated with a commercial of McDonald's led to an application showing the user's position and the closest McDonald's restaurant on a map.

2.2 Methods

A mosaic of different methods was applied in order to gather information from different perspectives and in different contexts. The most important ones were questionnaires, focus groups at the start and at the end of the trials, log file analyses and diary studies. A daily reminder to complete the online diary was sent to the participants at different times of the day via SMS. A link to the online diary was "bookmarked" on each device given to the trial participants. The online diary included open and closed-ended questions.

3 Results

Our analysis produced both quantitative and qualitative results upon which final recommendations were based to enhance the user experience of the decoder interface and of the Xaudio applications. Four of them are listed in section 4.

Figure 2 shows on its left hand side the distribution of the visited Xaudio applications. It is no surprise that this picture mirrors approximately the distribution of the different shows within Radio Hit's programme. On the right hand side the figure shows the relevance of the applications as stated by the participants. We see that although the application "News" was frequently visited, it did not provide additional value to the participants. The application "Commercial" shows a reverse pattern: Infrequently visited but highly rated. These results were also confirmed by the qualitative analysis of the focus groups and of the diaries and by the result shown in Figure 3 (Rating of the information provided by the last visited application).

The analysis of all the different methods applied shows that we may categorise three applications as successful (Traffic, Music and Commercials) and two as unsuccessful. However, the reasons for the success of these three applications differ. As "Traffic" mostly was used as a pull service the other two applications were used as push services.

Users who listened to a song or to a commercial break were suddenly reminded that more information on that song or on that product or store could be interesting and then actually accessed the corresponding Xaudio application. The traffic information was needed independently from the radio programme currently broadcasted (pull service). Some of the test users went to the start page of the Xaudio applications and accessed the traffic information from there, whereas the rest of them had forgotten that this short cut existed but nonetheless wished to have access to this information.

Since the Xaudio application "Traffic" was mostly perceived as a pull-service, it did not matter that its content did not deviate from the content broadcasted via the radio.

The other applications of Xaudio ("Weather" and "News") could not provide this added value. They were just considered as repetitions of the radio programme. Furthermore, the participants stated that missing one of these pieces of information did not matter and did not affect their private or professional lives. Especially the content of these two Xaudio applications, therefore, should be improved, in order to make sure that all the applications offer either an optimal push- or pull-service-content.

Although, the participants stated during the focus groups that the information provided should be as location based as possible, the LBS that was developed for this trial was not used very often. Since most of the participants knew the location of the closest McDonald's restaurant the information received was of limited use for the participants. In order to provide a real added value the covered area of the trials either would have had to be larger or the participants would have had to know the area not as well as the participants of our trials (inhabitants of Ljubiljana) did.

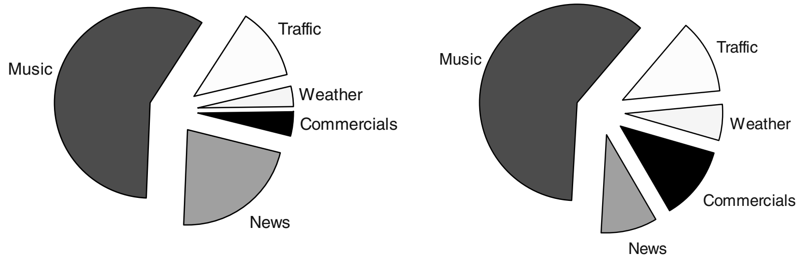


Fig. 2. Left: Distribution of the Xaudio applications visited by the trial participants (log files); Right: Most interesting Xaudio application since the last diary was completed as stated by the trial participants (diaries)

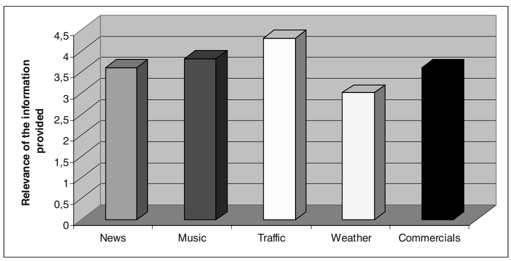


Fig. 3. Average relevance of the last visited Xaudio application as stated by the trial participants (diaries; 1: not relevant – 5: very relevant)

The field trials also showed that Xaudio's technical performance is still limited. Often users wanted to extract a particular watermark but failed for two reasons:

1. Participants located the device too far or too close to the speakers;
2. Participants used speakers, which were too "sophisticated" and which changed the acoustic composition and thereby deleted the watermark.

However, we have to notice that the trial participants on average accessed 4,5 Xaudio applications per day and that the whole system reached a SUS-score of 72,5. (The SUS (System Usability Scale) ranges from 1 – 100 [1].) Therefore, we could gather enough data to assess both the utility and the quality of the service and of the different Xaudio applications. Nonetheless, it seems that further research will be needed to make sure that this technology can attract users who are rather hostile to new technological developments and who are not as patient as participants of a field trial usually are.

4 Conclusions

The field trials showed that currently Xaudio's technology is not stable enough to make sure that all the listeners can receive the desired watermark independent from their current location (background noise) and independent from the speaker with

which they are listening to the radio. Nonetheless, the field trial uncovered lots of potentials to improve the decoder interface and the Xaudio applications (four of them are listed below). Their implementation will ensure that Xaudio can be used successfully in the future after the “cure” of Xaudio’s current technical “child diseases”.

Potentials to Improve Xaudio:

- More information displayed per watermark (e.g. singer or company name)
- Direct access to the Xaudio applications from the decoder interface independent from the show that is currently broadcasted
- More information at Xaudio’s start page (e.g. currently broadcasted song, latest commercial, etc.).
- Location based services should cover a broader area or should be designed for target groups such as tourists or business travellers

The field trials showed that Xaudio could provide added value to its users if it is either used as a push service (providing information for that users would or could not search efficiently) or as a door opener to a pull service that users may need in certain contexts. However, we believe that there are other scenarios that could be more promising to spread the publicity of Xaudio than the radio-scenario applied during this field trial. These other scenarios (music events, fairs, games at discos) would also allow controlling the conditions under which Xaudio is used. That means that the speaker system, as well as the loudness of the audio signal could be controlled to ensure an optimal watermark-extraction rate.

Acknowledgments. The field trial was conducted as a part of the Xaudio project, which was partly funded by the European Commission. We would like to thank all our partners for their inputs and for their comments. Especially we would like to thank Nikolaos Floratos (European Dynamics), Robert Kilbride-Newman (Central Research Laboratories) and Jani Krzic (Radio Hit).

References

1. Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In: Usability Evaluation in Industry. Taylor and Francis
2. Product features of the Xda II: http://www.myxda.com/XdaII/happenings/template/XdaIIProductInfo.vm (last visited 05/05/04)
3. Xaudio website: http://www.crl.co.uk/projects/xaudio (last visited 05/05/04)

O2 Active: Enhancing User Experience on Mobile Phones

Anat S. Amir

O2 UK, O2 Head office, Wellington Street, Slough, Berkshire
SL1 1YP, United Kingdom
anat.amir@o2.com
http://www.o2.com

Abstract. Mobile operators are providing customers a range of data services, extending beyond the traditional voice services. These aim to empower the handheld device or mobile phone as a central means of communication, information and socialization, changing the way people use and perceive them. The following paper describes the user interface concept of the O2 Active Menu. This unique mobile software application, launched in the UK in February 2004, provides an innovative interface on mobile phones, integrating between operator data services and the device functionality. It deals with the key user experience challenges of the mobile Internet environment, enhancing accessibility, presentation and usability.

1 Project Statement and Goals

The launch of i-mode and Vodafone Live! in Europe in 2002 heralded the return of the portal-based approach to mobile data content, following the poor adoption rate of WAP-based services in the preceding years [1], [7]. These mobile portals combine strong brands with content and applications optimized for particular handsets and networks, to deliver a more satisfying and consistent user experience. Page load duration has been mentioned repeatedly as a barrier preventing a wider acceptance of these services [3], [6].

The following paper describes how through user centered design, O2, a European mobile operator, created a unique user interface which bridges between the handheld device and online data services.

2 Challenge

The goal was to create and intuitive showcase for data service which would not be perceived as intrusive. The main design challenges in this project were to create an interface that will coexist with the device interface and existing WAP portal, so that users will be able to distinguish and understand the relation between the interfaces;

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 356–360, 2004.

and to create the best possible experience when accessing online content via the menu, launching the device browser, and returning from the browser to the menu

3 Design Process

A user centered design methodology was used to develop and define the requirements for the O2 Active operator menu. The Just-In-Time Usability Engineering approach was practiced to meet time to market while providing a high standard user experience [5]. This process included the iterative design of three prototypes that were presented to focus groups and evaluated during a number of formal usability testing sessions with customers in the United Kingdom, Germany and Ireland (50 participants in total). A trial version of the software was then distributed to 450 customers for feedback before launch.

4 User Interface

The first version of the O2 operator menu was designed for the Nokia series 60 Symbian devices [2], [4]. The graphic abilities of the series 60 allow developers to use the Symbian widgets and controls as well as to use graphics creating a new set of controls.

Several functional building blocks are available when implementing this type of software application, including: links to URLs; links to the device functions; links to other applications installed on the device; establishing voice calls; sending SMS/MMS; seamlessly updated text tickers; and screen sequences that can cater for purchasing of featured content or use data services through pre-defined wizards.

4.1 Accessing the Application

To increase discoverability, the menu is launched when the device is turned on. The menu is not displayed on top as the default application as this was perceived as intrusive by some customers, especially if they had personalized their phone display.

The menu is accessible via a number of routes: The device toggle key, the right idle screen soft key, and from a dedicated icon on the application grid. Assigning right soft key on the idle screen to this menu would ensure a high level of discoverability and accessibility. However, this option was not possible when the product was launched to market.

4.2 User Interface Concept

To enhance learnability three initial guidelines were set before designing the interface:

1. Interaction with the widgets and components in this interface should be consistent with the operating system.
2. Interacting with unique non-standard Symbian widgets (specifically using the arrow keys/joystick and softkeys) should be intuitive.
3. Key tasks such as keypad lock and dialing phone numbers using the keypad must be supported.

The interface is constructed of a two level hierarchy menu and consists of a number of 'wizards' facilitating sampling and purchase of content such as ringtones and wallpaper. The first hierarchy (O2 Menu) is split into two pages and consists of 12 categories, 6 on each page. Paging between these two screens is possible using the right and left device arrow keys (or joystick). The second level hierarchy screens are either presented in a list or a tabbed layout (two stacked lists). Three tool icons are presented at the bottom of the first hierarchy screens and a status bar was included at the top of all screens including several status indicators and the menu name. A dynamic news ticker is positioned at the bottom of the news category. (See Fig. 1 and 2).

Fig. 1. Home screen with all status indicators, and tooltip displayed when selecting tool icons

Fig. 2. Tab screen, List screen and Visual content. A globe is used as an indicator that selection will launch the browser and display online content. Icons are displayed besides menu items not in focus.

Findings from the usability studies held during the design and development process led to the definition of a number of design guidelines relating to tabs, widgets and layout.

Coexistence with Device Interface

- The layout and look and feel of the application should be visually distinct from the device application grid.
- The application status bar should be consistent with the device status bar.

Tab Design

- The visual appearance of tabs should reflect clearly what is the tab header and what are the tab content elements.
- A clear visual indicator should be provided to indicate tabs which are scrollable (the standard series 60 arrows at the bottom of the screen were not sufficient when displayed with tabs).
- When switching between tabs users expected the colours of the tabs not to change significantly (from blue to green for example).
- Both the right and the left arrow keys should be used to facilitate tab selection
- The right softkey displayed on 'Tab' screens should not be called 'Back' as it actually navigates 'Up' in the hierarchy and could cause confusion following switching tabs.

Other Guidelines

- Clear indicators should be displayed besides items that establish an online connection or a voice call when selected.

5 User Trial

Prior to launch a user trial was conducted with 450 participants in the UK. Participants had all used the O2 WAP portal once in the week prior to recruitment, owned a series 60 Nokia device and had both Internet and email access. The key objectives of the trial were to identify areas for further improvement and their reaction and perceptions towards this service.

Two surveys and two focus groups were conducted to collect feedback from participants during the trial. Participants identified the following key benefits:

- Improved presentation (mentioned by 33%).
- Useful shortcuts to specific functions (mentioned by 20%).
- Easier to access WAP (19%).
- Simpler to use than the WAP portal (17%).
- Allows to preview ringtones, wallpapers and other downloads before purchase (7%).
- Logical menu structure (14%).

The overall impression from the participants was very positive. More than 4 out of 5 users thought the O2 Active menu was an improvement when compared to the WAP

portal, which they used to access data service before the trial (due to several reasons, including improved usability, response time and attractive graphics). No issues relating to the user interface design were discovered, though some participants mentioned additional functionality they would like to see in this application. Therefore, O2 launched the service with the interface used during the trial, with the intention of further enhancements to future versions.

6 Summary

Page impressions on the O2 WAP portal increased by 25% (per unique visitor) within the first two months after launching the O2 Active Menu for series 60 devices. The user centered design process proved to be extremely efficient for the O2 Active Menu development project. It enabled release to market after a deep understanding of user perceptions, requirements and desired interaction.

References

1. Enders Analysis: UK Mobile User Survey, January 2004 [2004-03], http://www.endersaanalysis.com (2004)
2. Helle S., Jarnson J., and Koskinen T.,Takeout Menu: The Elements of a Nokia Mobile User Interface, in Lindholm C., Keinonen T, Kiljander H., Mobile Usability: How Nokia Changed the Face of the Mobile Phone (editors), McGraw -Hill (2003).
3. Nielsen Norman Group Report. WAP Usability Report, December 2000. http://www.nngroup.com (2000)
4. Nokia Corporation: Developer's Guide to Series 60 UI Category, Version 1.0 August, 2002. http://www.forumnokia.com (2002)
5. Reiman J., Just-In-Time Usability Engineering, in Lindholm C., Keinonen T, Kiljander H., Mobile Usability: How Nokia Changed the Face of the Mobile Phone (editors), McGraw -Hill (2003).
6. Roto, V., Kaikkonen, A.: Acceptable Download Times in the Mobile Internet. In Stephanidis, C. (ed.): Universal Access in HCI. Volume 4 of the Proceedings of HCI International 2003 (2003).
7. Yankee Group. Mobile Data Market Growth Remains on Course in Western Europe, January 22, 2004. http://www.yankeegroup.com (2004)

A Prototype for Remote Vehicle Diagnostics

Jonas Kuschel[1], Henrik Fagrell[2], and Carl Johan Andersson[3]

[1] IT University of Göteborg, Box 8718, 40275 Göteborg, Sweden
jonas.kuschel@ituniv.se
[2] Newmad Technologies AB, Thorildsgatan 7, 41125 Göteborg, Sweden
henrik.fagrell@newmad.se
[3] Volvo Parts AB, Product Support Development,
Dept 31410 ARH 7, 40530 Göteborg, Sweden
carl.johan.andersson@volvo.com

Abstract. The field of Remote Vehicle Diagnostics can be described as the remote management of vehicles equipped with electronic control systems. Despite the great potential that is ascribed Remote Vehicle Diagnostics there are few practical applications that address the needs of end-users. This paper asks how service mechanics remotely can get detailed vehicle data when the driver is concerned about the vehicle's behaviour, or the vehicle's internal control system detects an error. We describe a prototype that enables service mechanics to remotely receive notifications of vehicle diagnostics trouble codes, read real-time usage parameters, and periodic log parameters according to specified rules and filters. The paper concludes with a future outlook on how the architecture can support new kinds of services.

1 Introduction

In the automotive industry companies show an increasing interest in Remote Vehicle Diagnostics (RVD). RVD is the remote access, diagnosis and software update of vehicle systems. In 1998, Jameel *et al.* [2] predicted that new vehicles within five years would enable basic telematics services, such as sending status data and error reports via the Internet. While this has shown to be too optimistic, the interest in RVD among vehicle manufacturers, telematics service providers and end-customers is increasing [1].

There are a handful of commercial services, e.g., On Star or Volvo OnCall that all focus on consumer needs such as road assistance and guidance, but applications addressing the needs of service mechanics have not been in the spotlight.

We present an application prototype that aims to support service mechanics in identifying and solving problems remotely. The question we seek to answer is: *How can service mechanics remotely get detailed vehicle data when the driver is concerned about the vehicle's behaviour or the vehicle's internal control system has detected an error?* The application architecture underlying the prototype meets typical industrial requirements, e.g., cost per product unit, operation cost and scalability.

The model that has informed our design is primarily based on the results of an ethnographic field study reported by Kuschel & Ljungberg [3]. They conclude their work by proposing a decentralized approach to RVD. A part of the decentralized approach

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 361–365, 2004.

is to enhance after market mechanics with remote diagnostics access to the customers' vehicles. This perspective contrasts from the prevailing manufacturer-centric model of RVD where the local dealer mechanic plays a minor role or is totally removed.

The material was complemented with interviews and workshops with personnel from Volvo to further detail the requirements. The development was performed in co-operation with three master students using agile development methods. Finally, the prototype was evaluated as a proof-of-concept in a realistic environment.

2 Vehicle Electronics

We first introduce some technical concepts of vehicle diagnostics, since it is not a field normally addressed in HCI research. A modern vehicle is to a large extent controlled by computers, often called Electronic Control Units (ECUs). Figure 1 is an example of the electrical system of a Volvo truck. Several sensors are located all over the vehicle enabling the ECUs to monitor the status of onboard technology (e.g., fuel pressure) and the surrounding environment (e.g., barometric pressure). If a sensor value is outside the allowed range an ECU will signal an error code (sometimes called Diagnostic Trouble Code (DTC)). Error codes are categorized according to how serious the error is. A minor error would not be shown to the driver whereas a major one would force the vehicle to a stand-still position. Each ECU executes software and can thus be programmed to behave in different ways depending upon its application. For example, parameters such as injection ratio can be modified in order to control the performance, emission levels and fuel consumption.

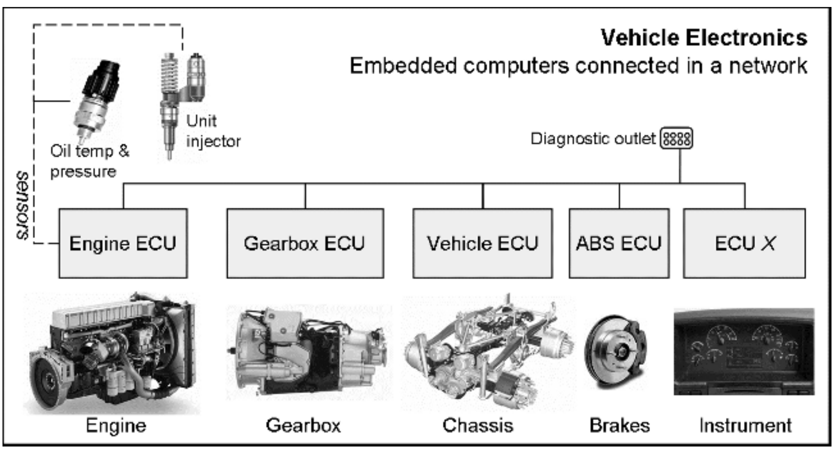


Fig. 1. The figure illustrates the electronic system of a modern vehicle.

Since a vehicle is such a complex piece of technology mechanics of today have to rely on computer programs to perform service. The diagnostic computer application used in the repair shop can be connected to the diagnostic outlet of the vehicle allowing the mechanic to read and reset error codes, read and set parameters, run scripted tests, and update the software of individual ECUs.

3 Meeting the Requirements

The prototype outlined in this paper aims to meet several requirements regarding mechanics' work practice, these are: (1) alerts when certain error codes occur; (2) reading run-time parameters; and (3) recording predefined parameters according to rules and filters.

By getting notifications about error codes the mechanic is able to start the diagnostic process and take action prior to the customer getting to the workshop. This would make the diagnostics process more efficient and thus improve customer satisfaction.

An important conclusion made by Kuschel & Ljungberg [3] is the fact that technicians define their jobs as identifying problems experienced by the customer, as opposed to technical problems as such. This requires mechanics to be able to analyze sets of run-time parameters since not all customer experienced problems are equal to error codes.

Both a problem description by the customer and diagnostic data stored in the ECUs are valuable clues that help the mechanic to define the problem. However, mechanics we have interviewed repeatedly point to the lack of relevant data in the time frame when a customer experiences a problem or an error code is set by the system. This becomes more evident regarding more difficult problems that require extended test driving and determining data defined by the mechanic. Hence, the prototype aims to enable mechanics to remotely define and record parameters according to rules and filters.

There are also several commercial and environmental constraints that have to be considered. Future telematics services must accord to the principle of low unit and maintenance costs.

4 The Prototype

The prototype builds upon a uniform remote communication module. It handles network breakdowns, roaming between networks and truncation of the data stream.

A PDA was used as onboard client because the platform allows for speedy development. In a commercial application the onboard functionality would be developed for a less sophisticated device (i.e., ECU) to reduce the hardware costs. The server is PC-based with a Java platform. The mechanic interacts with the system via a web browser (see figure 2). Needless to say, the communication with the vehicles must be wireless.

Three services have been implemented in the communication module; an error code notification service, a service for run-time parameter reading, and a service for recording of predefined parameters.

The notification service informs the mechanic about error codes recently occurred. Only changes in the ECU states are sent to the server. Accordingly, the server will always keep the latest state about the vehicle and the mechanic is able to get data about a vehicle's state immediately without waiting. While connected to the server a notification is sent in order to get the mechanic's attention (see figure 2, note 1).

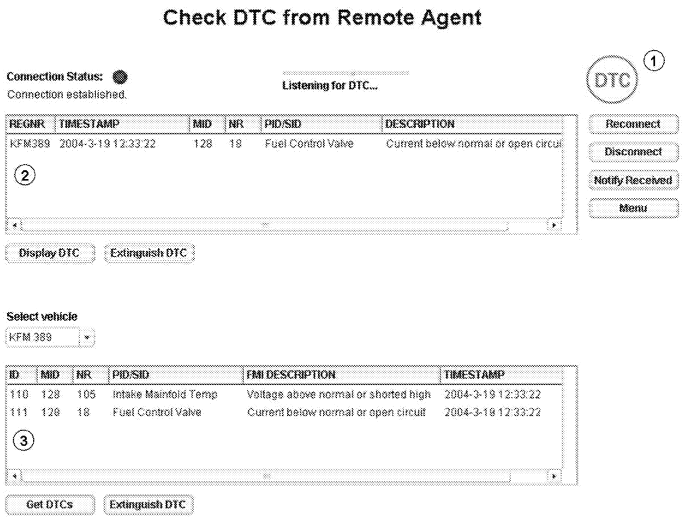


Fig. 2. A screenshot of the web interface, with (1) a notification about a recently set error code, (2) data about the error code, and (3) data about previous error codes.

Fig. 3. The setting of the proof-of-concept evaluation.

The service for reading run-time parameters enables the mechanic to remotely read parameters, e.g., boost pressure. This is a rather straightforward service that, by getting a request from the web client, requests a parameter from the ECU chosen and sends the data back to the mechanic.

The third service enables mechanics to log parameters according to five different settings; parameter, total time, interval, frequency, and rules. The mechanic can select different parameters and define how long each parameter should be read. An interval scheduling how often data is sent to the server and the frequency of reading the parameter from the ECU are further settings to be defined. Finally, rules can be set that define, for instance, a data range between which the parameter should be logged. The log service enables mechanics to conduct more profound analysis remotely. Up until now, so-called flight recorders, storing data in a hardware unit, have to be used for this purpose. Our prototype enables mechanics to continuously analyse the data and, most important, change the settings during operation.

Summarizing all three services they operate according to cost effectiveness and mobile computing constraints. Most important, our ambition has been to develop an architecture that enables a smooth interaction with the application. Error codes or re-

quested data are continuously sent to the server and cached in a database, thus offering a quick access and only notifying a mechanic when any changes occur.

The proof-of-concept evaluation was conducted at a test drive track (figure 3). The setting of the evaluation was (1) a laptop connected to the Internet and running a Web browser; (2) a PDA and an interface jacket connected to the diagnostics outlet of the Volvo FH 12 (3). The PDA was connected to the Internet via Bluetooth to a GPRS phone giving us wireless access to the truck when it was on the road (4).

All of the services of the prototype were successfully performed. Despite low network coverage of the test track site all data could be transferred via GPRS. This indicates the level of data efficiency we were able to achieve.

5 Discussion

In this paper we have addressed the issues of how service mechanics remotely can be provided detailed vehicle data when the driver is concerned about the vehicle's behaviour or the vehicle's internal control system has found an error. The proof-of-concept prototype has been tested under realistic conditions with promising results. As a next step a new service for remote parameter setting and ECU software updates is going to be introduced. We also plan to evaluate the complete systems on professional mechanics. In doing this, we believe that there is a great potential to find new requirements on how mechanics want to remotely interact with the vehicles and the customers in order to complement the diagnostic data.

References

[1] Gartner G2, Telematics Industry Outlook: Think Outside the Vehicle, RPT-0902-0163 Gartner G2, 2002.

[2] Jameel A., Stuempfle M., Jiang D. and Fuchs A. [1998], Web on Wheels: Toward Internet-enabled Cars, *IEEE Computer* 31(1), 69-76.

[3] Kuschel J. and Ljungberg F. [2004], Decentralized Remote Diagnostics: A Study of Diagnostics in the Marine Industry, to be published in *Proceedings of HCI 2004*, Leeds, UK.

Adaptive Portal Aggregation for Pervasive Client Devices

Ming Li[1] and Marion Blount[2]

[1] Computer Science Department, Dartmouth College, Hanover, NH 03755, USA
ming.li@dartmouth.edu
[2] IBM T. J. Watson Research Lab, 19 Skyline Drive, Hawthorne, NY 10532, USA
mlblount@us.ibm.com

Abstract. Accessing portals using pervasive client devices, such as PDAs, smart phones, etc., has become very important to mobile users. The characteristics of pervasive devices, such as small form factor, screen geometry diversity, low processing power, weak network connectivity, etc., impose many challenges for accessing portals on the pervasive devices. A major overlooked challenge is the difference in user attention on a desktop system and the user attention on a pervasive device. In this paper, we propose an adaptive portal aggregation framework to minimize the user attention demands for accessing portals on pervasive devices. We propose two specific techniques and a general approach that uses client context information for adaptive portal aggregation for pervasive client devices. We have tested these approaches in an embedded portal that runs on a PDA.

1 Introduction

Portals have become one of the most important ways of delivering applications to users. A portal is conceptually an entrance to a collection of web-based applications. Portal applications are packaged and presented as web applications called "portlets." The three-tier web model [2] makes it possible for multiple users to share a portlet installed on a portal. The total cost of the ownership of the portal is much lower than the cost of having the portlet application replicated at each pervasive device. To update the portlet application, one only needs to update the portlet on the portal. System management issues are eased for both the enterprise and the end users.

Pervasive devices such as PDAs and smart phones are being used more and more by mobile professionals. Most of the devices have web browsers and wireless connectivity as standard features or options. An increasing percentage of the applications they access are portal applications. Hence, supporting good portal access from pervasive client devices is vital.

Compared with accessing portals on desktop systems, there are several prominent challenges for accessing portals on pervasive devices [3]. Those challenges directly map to the key design concerns of multi-device portal developers.

- *The on-the-road factor*

Desktop users usually apply a substantial amount of their attention to interaction with desktop applications. The amount of information on the screen, the location of the cursor, and the placement of hot links are factors that demand attention. When mobile, users sometimes use applications on the pervasive device as a part of another

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 366–372, 2004.

task like making a phone call or trying to navigate to a destination. Scrolling and pull down menus are distractions on pervasive devices. Minimizing the distractions of using the mobile device is very important.

- *Small screen factor*

For a typical desktop view of a portal page, many portlets can be simultaneously laid out in a table format on a screen. It is easy to navigate among the portlets. It is a great challenge to present multiple portlets on a screen of a pervasive device without excessive scrolling.

- *Geometry variety*

The geometry of the pervasive client devices varies quite a bit. For desktop PCs, full-size landscape-oriented screens are the norm. Tablet PCs usually have full-size portrait-oriented screens. For PDAs, palm-size portrait-oriented screens are most popular, but landscape-oriented ones, such as Zaurus, also exist. Pervasive devices, such as smart phones or smart watches have smaller screens with different orientations or shapes. Fig. 1 shows some examples of the varied screen geometries of pervasive devices.

Fig. 1. Different Geometries of Mobile Devices

Fig. 2. Indexical Portal Front Page

- *Limited computing resources*

Processing power and network connectivity associated with pervasive devices are usually less than those on desktop computers. Some portlet application features may not be useful on pervasive devices.

While each portlet generates its view markup, the portal aggregator controls the placement of the markup on the screen and the navigation among the portlets. In this paper we propose extensions to the portal aggregator that allows it to enhance the experience of accessing a portal with a pervasive device. The goal is to minimize the attention and steps a user must devote in accessing a portal with a pervasive device. The work derives from the experiences the authors had with an embedded portal. Key elements of the proposals have been implemented in an embedded portal.

We first briefly review the state of the art of portal aggregation. We next propose and explain adaptive portal aggregation. We end with conclusions and future work.

2 Portal Aggregation for Pervasive Devices: State of the Art

One can make some key observations about the state-of-the-art portal support for pervasive devices. Portals use the user-agent field of the http header to determine the type of browser, hence the device. The portal then returns markup designed for PCs or markup designed for PDAs. We us the term *front page* to refer to the page that presents the user the choice of portlets to access. For pervasive devices, the front pages use icons or string names/descriptors which are hot links to the portlets (see Fig. 2).

Current front page design requires the user to proactively click the icon or link and load the individual portlet to check the status or new events. Such proactive actions demand a lot of user attention. If there is no new email message at the moment, users do not need to interact with the email portlet. In addition, such design could potentially waste processing power or network bandwidth if, say, a user connected with the mail server and loaded the whole inbox only to find no new messages.

The rigid representation of portlets in icons or links on a front page may in some cases require excessive scrolling effort to view all the included portlets. This partly defeats the purpose of having a front page. A front page is meant to provide an index to portlets in a portal with only a glance.

3 Adaptive Portal Aggregation

In this section we investigate ways of augmenting the portal aggregator with additional information which it uses to adapt its portal page presentation in ways that enhance the user's portal access experience. In particular, we deal with the challenges specified in the introduction. We call the approach *adaptive portal aggregation*.

3.1 Summary View Mode

We have designed and implemented a new portlet mode, summary view mode, to exploit the small display of pervasive devices. In this mode, a portlet shows the key status information and only use a fraction of the screen area it would use for a full portlet view. The key status information presumably includes the events and information that are most interesting to the user or mostly likely to trigger user interaction with the portlet. Fig. 3 illustrates the summary view mode for two portlets we implemented. Normally, only one portlet can be viewed on the iPAQ PDA, but, in summary view mode, the key events and information of two portlets can be viewed. The user can determine if there are any new emails with a quick glance of the portlet. Thus, we minimize the user attention needed to gather key information from the portlet.

Just as a portlet developer must implement a portlet view for edit, view, and help modes, a view must also be developed for summary view mode. The portal must also support the new mode. We were able to add support for summary view mode in our pervasive device portal very easily.

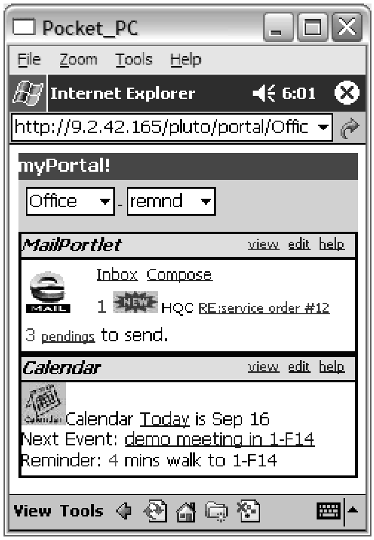


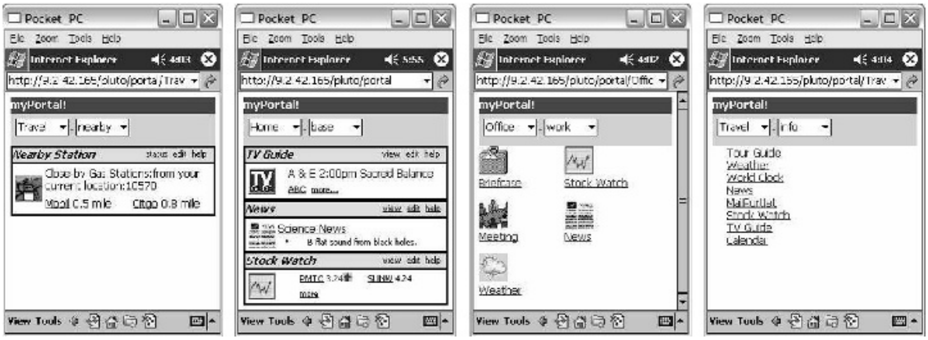
Fig. 3. Front Portal Page in Summary View Mode

The information and events to display in summary view mode are generally obvious for a given portlet. In the case of the email portlet, the status of the Out Box and the number of unread emails in the In Box are key data. For the calendar portlet, information on the next event is key data.

3.2 Front Page Adaptation

We use the term front page to refer to any portal web page that gives the user a set of portlets to select. While the front page on a PC screen can give the user multiple portlets in view (full) mode, the front page view on pervasive devices can display at most one portlet in view mode. Front page views on pervasive devices almost always involve scrolling, especially vertical scrolling. The goal of front page adaptation is to minimize vertical scrolling and to maximize the amount of information on the viewable area of the front page. To this end, the portal aggregator needs to be aware of the size and orientation of the device screen and how many portlets are included in a portal page. Then it picks the most appropriate representation of portlets to fit into a screen (see Fig. 4).

Based the number of portlets on the front page, and the value of threshold values, the portal aggregator determines how to display the front page. In the example of Figure 4, a single portlet in the front page results in displaying the portlet in view mode. If there are between 2 and 3 portlets in the front page, the summary mode view is selected for those portlets. If there are between 4 and 6 portlets, the icon view is used, and more than 6 results in the list view. This adaptation can be made highly customizable; a user can change the threshold values to reflect his own preferences. The portal aggregator is using information available to it to minimize user distractions, like scrolling, and to enhance the user experience of accessing a portal using a pervasive device.



a) single portlet view b) summary mode c) icon view d) list view

Fig. 4. Front Page Adaptation Techniques

3.3 General Framework for Adaptive Portal Aggregation

The screen size and orientation of the pervasive device and the number of portlets on the front page is an example of the portal aggregator using local client context information to enhance the user experience of accessing the portal on the pervasive device. This can be generalized to having local client context information passed on the aggregator to make it even more adaptive. Fig. 5 shows a general context-aware portal aggregation framework. The Context Collector is a subsystem that runs on the pervasive device and has the task of being the collection entity for local device context. There are a number of peripherals and sensors that can be attached to pervasive device today; there will be even more in the future. With a GPS attachment, the device will be able to discern home location from work location. Light meter attachment can indicate the brightness of the light in the vicinity of the pervasive

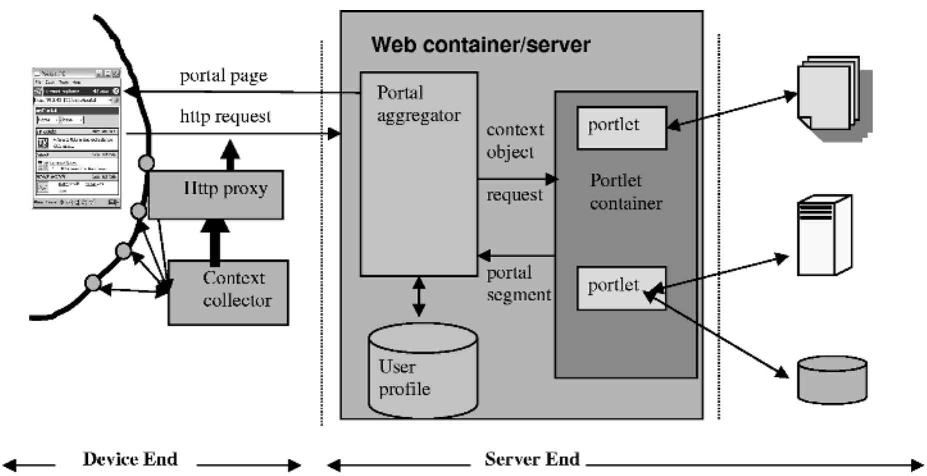


Fig. 5. General Framework for Adaptive Portal Aggregation

device. Local device context is passed to the portal by an HTTP proxy which sends the context information in special header fields of HTTP requests. The portal aggregator receives the HTTP request, strips the context related headers, and updates a repository of device context data. The portal aggregator uses this context data when it aggregates the portlet fragments and formulates a response to the client device.

If the HTTP request originated from a link in a portlet the final destination of the HTTP request is a portlet. So, we have a technique for passing client context information on the portlet. The context information can be passed on the header fields or in a context object. A portlet does not have to use the context information. A number of interesting adaptations now become possible. If the portlet can determine from the local context information that the pervasive device is in view of other persons, it can generate the public view of the portlet. For example, a banking portlet may gray out bank balances. The portlet can generate view markup that is easier to view in bright light.

4 Conclusions and Future Work

The work described in this paper is an extension of some earlier work on an embedded portal [1]. We designed and implemented an embedded portal that is based on the Reference Implementation of the JSR 168 Portlet Specification [4]. The embedded portal, implemented in Java, runs on OTI's J9 JVM [5]. The embedded portal runs on the Compaq iPAQ and Windows PCs.

Based on our experiences with the embedded portal, we explored ways of enhancing the user's experience of accessing a portal with a pervasive client device. At the start, accessing a portal with a pervasive device took the same amount of user attention and proactive steps (clicks) as accessing a portal on a PC screen. Our goal was to minimize the user attention and steps for accessing a portal on the pervasive device. We proposed, designed, and implemented support for summary view mode in the embedded portal and JSR 168 portlets. With a summary view of a portlet, the users see all the key status information of the portlet without proactively loading each individual portlet. Our front page design and implementation adaptively chooses suitable layout and representations of portlets to minimize the scrolling effort to view all the portlets on a front page.

After designing and implementing the aforementioned adaptations, we realized that there was a need for a general framework to gather and pass local pervasive device context information to the portal aggregator. The key to adaptive portal aggregation is making the portal more aware of the additional client-device side context information, namely 1) the physical characteristics of the device, such as the size, shape of the screen, supported software on the device, and 2) the context information that could be sensed by the device itself, such as the location, nearby people, etc. We expect to see more and more interesting context information that can be obtained by the devices through embedded sensors on the devices or by attachments to the devices.

TGL Micro [7] is developing a web service framework for collecting client information from the mobile devices. This is an alternative way for a portal to gather local context related to a pervasive device. The portal can subscribe to get updates of changes in the local context of a pervasive device.

The next step is to devise and perform a user study to illustrate the advantages of these adaptive techniques for accessing portal with a pervasive client device.

References

1. Marion Blount, Veronique Perret, Danny Yeh, Apratim Purakayastha , Michael Moser, Yann Duponchel, Daniela Bourges-Waldegg, Marcel Graf: Managed Portal Appliance: An Experiment in Extending the Reach of Web Applications. In IEEE International Conference on Mobile Data Management, Berkeley, CA, January 2004
2. Paul Ashley, Sridhar Muppidi, Mark Vandenwauver: Building a pervasive computing portal. http://www-106.ibm.com/devloperworks/library/it-0701art13/?dwzone=ibm
3. D.M. Russell, P.R.Maglio, R.Dordick, C. Neti: Dealing with ghosts: Managing the user experience of autonomic computing. In IBM Systems Journal Vol. 42 No. 1 2003
4. JSR 168 specification http://www.jcp.org/jsr/detail/168.jsp
5. IBM Service Management Framework http://www-306.ibm.com/software/wireless/smf/
6. Stitch together views to facilitate construction of Web applications http://www-106.ibm.com/developerworks/java/library/j-strutstiles.html?loc=j
7. TGL Microsystems http://www.tglmicro.com/

Supporting Mobile Applications with Real-Time Visualisation of GPS Availability

Anthony Steed

Department of Computer Science, University College London, Gower St, London, WC1E 6BT, United Kingdom
A.Steed@cs.ucl.ac.uk
http://www.cs.ucl.ac.uk/staff/A.Steed/

Abstract. Many mobile applications rely on the Global Positioning System (GPS) to provide position and location information. However, there are many problems with using GPS in urban environments due to the variable nature of GPS's accuracy and availability. This paper introduces a simple tool that visualises the current state of GPS availability in real-time. This tool can be used for scenario planning for certain types of mobile applications and as aid for analysis of location logs.

1 Introduction

Many mobile applications require positioning information [1], and those that operate outdoors often use the Global Positioning System (GPS) system. Applications that rely on GPS range from widely available guiding and mapping applications, through location-based services [2] to augmented-reality style presentations [3].

GPS technology is a rapidly moving field. However a consumer-grade GPS unit typically needs to be able to detect the signal of three or more of the GPS satellites in order to be able to generate a position. At the time of writing there are thirty-one GPS satellites in orbit that provide good coverage given an open sky. However, in urban environments, the skyline can have a very significant elevation. This restricts the amount of sky that can be seen and reduces the likelihood of being able to see the requisite number of satellites. A second complicating factor is that the satellites are in non-stationary orbits, so even if a GPS unit is in a static position, the GPS availability will change over time.

In this paper we introduce a tool, *satview*, which visualises the current likely availability of GPS coverage. The tool takes a 3D model of the local environment, and the satellites positions. In real-time, the tool shows where on the ground plane one would likely be able to see three or more satellites and thus reliably get a position fix. We developed this tool in response to two mobile application scenarios in the EQUATOR project. We will briefly discuss how we have used or plan to use satview to improve the effectiveness of users in these scenarios.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 373–377, 2004.

2 Visualising GPS Availability

GPS units usually report navigation information using the NMEA 0183 communications standard [4]. Aside from messages that give position estimates, NMEA 0183 supports a message that describes the satellites in view. This message gives the azimuth and elevation of up to 12 satellites that have been calculated to be in the sky above the unit, along with a signal strength for each.

If we have a 3D model of the area near the unit, we can easily visualise the visibility of any one satellite using a graphical shadow algorithm. If we consider that satellite to be a light source, then everywhere that is in shadow will be hidden to the satellite. Fig. 1. shows a 3D model and the areas of the ground that would be visible and invisible to one satellite. Our model was generated using the automatic process described in [5] which uses commonly available vector map data. A review of methods for generating urban models can be found in [6].

a) b)

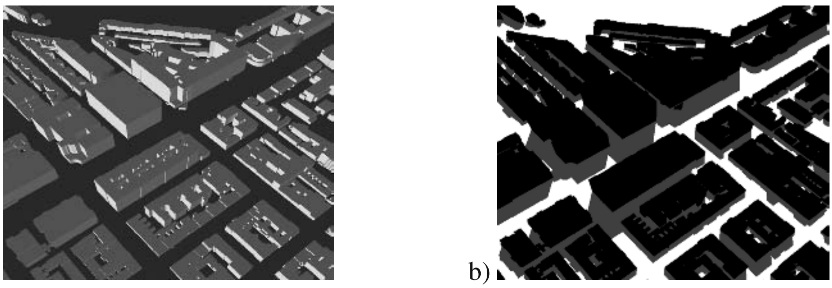


Fig. 1. a) Simple 3D block model of an area near Baker St in London, UK. b) Visualisation of the visibility of a single satellite in that area. White regions on the ground can see the satellite, grey cannot. The buildings are drawn in black

In our scenario, because we are interested in visibility on the ground plane, an algorithm such as the "fake-shadow" algorithm is sufficient [7]. This algorithm works by projecting all the scene geometry on to the ground. In a graphics API such as OpenGL [8] this can be done with a by rendering the buildings after pushing an appropriate projection matrix onto the modelling matrix stack.

In order to combine multiple shadows and find regions that satisfy various conditions on visibility, we use the stencil buffer that is available on most 3D graphics accelerators. The stencil buffer is a set of bit planes into which values can be written (refer to [8] for a detailed description). Like the depth buffer, the stencil buffer is not directly shown on the screen, rather arbitrary binary tests can be made against the stencil bit planes when filling pixels in other buffers.

For each frame we proceed as follows:

1. Clear the colour, depth and stencil buffers
2. Draw the city model and ground plane into colour and depth buffers.
3. Disable drawing to colour and depth buffer but leave depth testing enabled.
4. For all *i,* draw the fake shadows from satellite *i* into bit-plane *i* of the stencil buffer.

5. Read the stencil buffer back
6. Enable colour drawing. Disable depth testing.
7. For every pixel on the screen, if the value in the stencil buffer indicates that our visibility condition is met, then plot that screen pixel in white.

For Step 7 we have to interpret the bit mask returned in the stencil buffer as a visibility condition. For example, if there are N satellites, and we are interested in every point that sees three or more satellites, we need to plot all points for which no more than N-3 bits are set in the stencil buffer. Given that a maximum of twelve satellites are visible this can be easily implemented with a lookup table. Other visibility conditions can also be encoded in a similar manner. A minor implementation issue is that graphics cards vary in the number of bit planes they support in the stencil buffer. Eight is a typical value. This implies that the stencil buffer might need to be read and cleared multiple times.

Fig. 2. shows the visualisation tool itself.

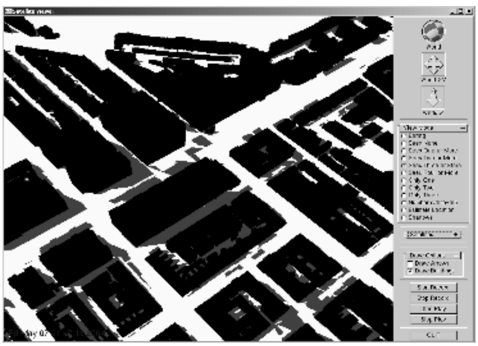


Fig. 2. The satview tool. All areas that can see three or more satellites are shown in white

3 Scenarios of Use

The development of satview tool was motivated by two scenarios. The first was the experience of colleagues who used GPS as the positioning technology in the Can You See Me Now? (CYSMN) game event [9]. In CYSMN runners in the real world were tracked using GPS. They had to “catch” online players who were navigating about a map of the same area. A catch occurred when a runner got a GPS position fix in the vicinity of the position of an online player. The runners would often experience *black spots*: areas where GPS position fixes were so inaccurate that they would be unlikely to catch nearby online players. This caused frustration for the runners. Inaccuracy would also cause frustration for the online players because they could get caught by one of the inaccurate position reports from a runner.

Designs for future interfaces for such events have proposed two strategies for dealing GPS uncertainty [10]. The first uses historical logs of GPS position fixes to identify likely black spots. The second uses satview to predict current black spots. The

latter is only useful if it can be conveyed to the runners in real-time, either by presenting maps on the PDA, or relaying instruction via an operator.

The second scenario is concerned with the collection of logs of pollution levels. On the Advanced Grid Interfaces for Environmental e-science in the Lab and in the Field project we are investigating the use of GPS-tracked pollution monitors to make dense maps of pollution [5],[11]. In one study we are attempting to make a map of an area of London that is a case study for the Dapple project [12]. We are using satview to plan which regions to map in particular sessions, by noting whether east-west or north-south roads are more favourable and whether there are enough high satellites to survey narrow streets. Fig. 3. shows how GPS availability was predicted to vary over an hour at one road junction.

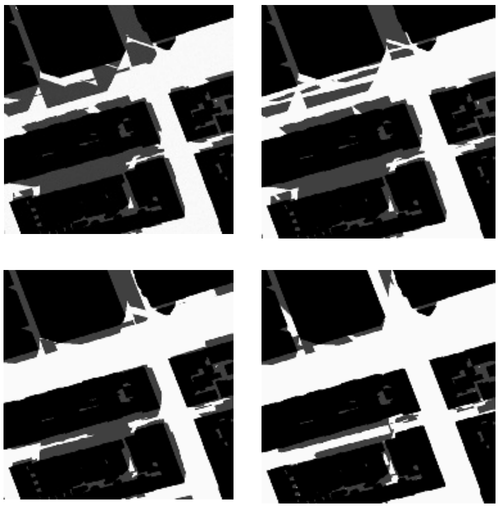


Fig. 3. Predicted satellite availability changing over an hour

4 Discussion

The satview tool provides a simple way of visualising GPS availability. At the very least it provides a useful tool for explaining why users' experience of GPS positioning is as varied as it is. It also provides a tool that can be used in certain scenarios to plan when to visit a region.

This is a work in progress, and in the near future we will add the capability to extrapolate satellite paths to predict availability over the next few hours. We are also looking at the requirements for detailed modelling of building heights. Our model of Baker St uses only crudely estimated heights but we do have detailed height models of other parts of London [5]. Note that with this tool we can't model the effects of reflected signals, nor of signal diffraction.

In the longer term, the tool might provide a method of improving GPS accuracy by exploiting knowledge about visibility and invisibility of satellites to provide a constraint on the position fix. It is worth noting that accuracy and availability issues are

being vigorously attacked in the navigation community. In the longer term systems such as Galileo that are complementary to GPS will be launched and these will improve accuracy in urban canyons [13]. However we expect that a tool like satview would still be useful to detect potential black spots or regions of differing accuracy due to different numbers of satellites being available.

Acknowledgements. This work was supported by, EQUATOR Interdisciplinary Research Collaboration (EPSRC Grant GR/N15986/01) and Advanced Grid Interfaces for Environmental e-science in the Lab and in the Field (EPSRC Grant GR/R81985/01). Thanks to Richard Milton for generating the models of the Baker St area. Movies of satview are available from the author's web page. The software itself is also available on request.

References

1. Hightower, J., Borriello, G.: Location Systems for Ubiquitous Computing. Computer, 34(8), August (2001) 57-66
2. Cheverst, K., Davies, N., Mitchell, K., Friday, A., Efstratiou, C.: Developing a Contextaware Electronic Tourist Guide: Some Issues and Experiences. Proceedings of CHI 2000, The Hague, The Netherlands, 17-24 March (2000).
3. Feiner, S., MacIntyre, B., Höllerer, T., Webster, T.: A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. Proceedings First IEEE Int. Symp. on Wearable Computers, October 13-14, Cambridge, MA (1997)
4. National Marine Electronics Association, NMEA-0183 Standard.
5. Steed, A., Spinello, S., Croxford, B., Milton, R.: Data Visualization within Urban Models. Theory and Practice of Computer Graphics 2004, 8-10 June, University of Bournemouth, UK (2004)
6. Shiode, N.: 3D Urban Models: Recent Developments in the Digital Modelling of Urban Environments in Three-Dimensions. GeoJournal 52 (3), 263-269
7. Blinn, J.: Me and My (Fake) Shadow. IEEE Computer Graphics and Applications, 8(1), (1988), 82-86
8. Shreiner, D., Woo, M., Neider, J., Davis, T.: OpenGL Programming Guide 4th Edition. Addison Wesley (2004)
9. Benford, S., Anastasi, R, Flintham, M., Drozd, A., Crabtree, A., Greenhalgh, C., Tandavanitj, N., Adams, M., Row-Farr, J.: Coping with uncertainty in a location-based game. IEEE Pervasive Computing, September (2003) 34-41
10. Crabtree, A., Benford, S., Rodden, T., Greenhalgh, C., Flintham, M., Anastasi, R., Drozd, A., Adams, M., Row-Farr, J., Tandavanitj, N. and Steed, A.: Orchestrating a mixed reality game 'on the ground'. Proceedings of the 2004 CHI Conference on Human Factors in Computing Systems, April 24-29, Vienna (2004).
11. Advanced Grid Interfaces for Environmental e-science in the Lab and in the Field (Urban Pollution), http://www.cs.ucl.ac.uk/research/vr/Projects/envesci/, accessed 4 May 2004.
12. Dapple, http://www.dapple.org.uk, accessed 4 May 2004.
13. O'Donnell, M., Watson, T., Fisher, J., Simpson, S., Brodin, G., Bryant, E., Walsh, D.: Galileo Performance: GPS Interoperability and Discriminators for Urban and Indoor Environments. GPS World, June (2003).

Bringing the Office to the Stables

Stefan Wagner

EIT-Department,
University College of Aarhus,
Dalgas Avenue 2,
DK-8000 Aarhus C, Denmark

Abstract. This short paper presents partial results from a research project aiming at uncovering the potential for introducing mobile computing support in the Danish Agricultural sector. Using a commercial research project as the staging point, several prototypes are developed and presented to farmers in a user study, combined with an on-site field study aimed at gathering information on the present usage of computer application support, as well as uncovering potential relevant usages for existing mobile technology devices, like PDAs and cell phones. The main lesson learned from this research project is not to replace existing tools and technologies with new mobile technology at any cost, but rather utilize these new devices to supplement and augment the existing infrastructure already in place.

1 Introduction

Maersk Data Food & Agro, a major Danish software engineering company specializing in software solutions for the Agricultural sector, performed an initial survey in 1999 of the potential for mobile computing support for farmers. This resulted in a white paper, recommending the start up of a series of prototype projects, trying to "convert" the company's existing agricultural related applications from the stationary personal computer platform, to mobile devices like personal digital assistants (PDA) and cell phones.

Based on the initial findings of this commercial project, a research project was established to investigate whether the introduction of mobile computing solutions to the Agricultural sector might be feasible or not.

This paper begins with a motivating discussion concerning the need for mobility in the Agricultural sector and the challenges associated with implementing it, as a basis for a further investigation of the problem area. After describing a qualitative user study of the work being done in the stables and offices of Danish farmers, consisting of field studies and interviews with a total of fourteen farmers and farmer's assistants, the paper analyses the observations made, and reflect upon them.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 378–382, 2004.

2 Motivation and Background

Most industrial sectors utilizing computing support are today employing technologies and applications dominated by the personal computing paradigm. This also holds true in the Danish Agricultural sector, where the personal computer has become an indispensable tool for planning, managing and reporting in all aspects of the work. It is almost inconceivable that e.g. a Danish Pig Farmer could manage his production, often up to 10,000 or 20,000 pigs per year, without the use of these planning tools, also known as management tools. The high wages in Denmark do not allow for employing more than 2 to 4 assistants, which calls for a high level of efficiency in order to handle production levels of these sizes. In fact, there are several other technologies (besides the management tools) that are vital to most pig farmers, including automatic feeding and drinking-water systems amongst others, but this is a huge area in itself, and will not be covered by this paper.

The use of stationary PC-based applications (the *management tools*) for storing all data concerning the animal production might seem a bit odd, considering that the data is actually needed, not in the office where most farmers keep their PCs, but rather in the stables, where the decisions are to be made. Indeed one could argue that the PC has been chosen as the platform of choice, not because of its suitability for solving the computing needs of the farmers, but rather because it historically has been the only technology available.

This does present at least one important issue: the distance from where the information is available, to where it is needed. Several researchers have already discussed this topic including Marc Weiser [8, 9], Donald Norman [6] and Poul Dourish [3]. In the case of the pig farmer, who only spends a fraction of his working day in the office, using a personal computer is not the optimal solution. Amongst other, he has no access to the PC-based data while being in the stables, where the data is needed and most of his work takes place. He is forced to produce paper printouts in advance, making handwritten scribbles for future entry on the PC, or simply just disregard the data altogether, resulting in a potential loss of productivity. Likewise, in related work studies of mobile and nomadic users, including the field studies performed by Nielsen & Søndergaard at a wastewater treatment plant [5], and by Bardram et al. at a Danish hospital ward [1], analogous observations support this.

Considering this, it does actually seem quite apparent that farmers might be candidates for mobile computing technology, considering the highly mobile nature of their workplace

3 The User Study

Several Danish farmers and their assistants totalling fourteen individuals participated in a user study to gather data on the subject under scrutiny.

3.1 Research Methods

The user study was split into two phases: *Qualitative interviews* & *Field studies*. First, using Steinar Kvale's model for performing a *Qualitative Interview* [4], the participating farmers were introduced to the ideas and possibilities of the mobile computing paradigm, as well as several mobile computing prototypes (see next section) for clarifying the technical possibilities available. This was followed by a structured discussion of ideas and general views, which in term was recorded for further analysis.

Second, a field study was completed at each farm, with a researcher following the everyday work tasks at the stables and in the offices of the farmers. The field studies were all based on the *Contextual Inquiry* method, as discussed by Beyer & Holzblatt [2] and Väänänen-Vainio-Mattila & Ruuska [7], modified to suit the needs of this particular project. The goal of the field studies was to document most of the typical work tasks, especially those requiring access to PC-based data.

3.2 The Farmers

Most of the participants in the study were highly professional Danish farmers and farmer's assistants. This implies several years of agricultural training at a public agricultural school as well as an apprentice at a farm. The former including formal training in using computers and agricultural related software. Only one participant of the user study was a "*part-time farmer*", with a regular job not related to the Agricultural sector, and no formal training in farming.

Most farmers were already confident with using cell phones and a PC, both in their spare time and professionally, but some of the assistants were not intimate with the use of computers, and were reluctant to use one, while all were comfortable with cell phones. In fact many farmers had come to rely on the cell phone for communication and coordination with their employees (the assistants). As most Danish pig farms are spread over several physical locations, this makes them especially useful.

3.3 The Prototypes

Several prototypes were developed for gaining experience with the technical possibilities of the mobile technologies, and for providing a foundation for the farmers to evaluate the potential of these. Only a subset of these prototypes is presented in the following. For the cell phone devices, prototypes were developed using *WAP & Wireless Markup Language* (WML) technology, as well as *Java 2 Mobile Edition* (J2ME) and *Windows CE embedded C+,* which was also used for the PDA versions of the prototypes (the Pocket PC platform). The prototypes span from being "mobile versions" of existing applications, e.g. pig management applications like the *BEDRIFTLOESNING* or *WinSvin*, to more novel uses like cell phone based camera-surveillance of animals.

3.4 Observations from the User Study

The interviews and field studies confirmed that all farmers spend most of their time away from the office, without access to the PC and the management applications (e.g. BEDRIFTLOESNINGEN or WinSvin).

Their work is of a highly mobile nature, and the use of IT is secondary to almost all work tasks, meaning that the computer is not needed during a normal workday. Most of the work tasks were however highly dependent upon a printed paper list for information on what to do next (a *planning list*). For example, there would be a list indicating which sows (female pigs) were to be inseminated (impregnated) at a specific date (e.g. sow number 1040, 1011, 923 … at May 12th). Without these lists it would be near impossible to keep track of the current status of each pig. Likewise all other aspects of the life cycle of the animals are precisely planned in detail by the management application.

After having performed the scheduled tasks, the paper lists are updated by hand (pencil), and by the end of the week, the lists are collected and updated at the office PC. Subsequently new lists for the following week are printed, and taken to the stables. This entire process takes no more than 20-40 minutes. This includes follow-up studies of production reports, indicating whether the production is running efficiently or not.

Although most farmers recognize the paradox of them being highly nomadic workers and still being forced to use the office based stationary PC, they point out that some work tasks demands the wide screen area of the PC, and that these tasks actually are better suited for the office. Thus, the user study does not indicate any potential for replacing the PC with mobile devices all together. With regard to the paper-based planning lists, it was suggested that mobile devices might replace the paper-based media, thus eliminating the need for transferring data from paper to PC at a later point. The main observations made were, however, that the farmers themselves were quite satisfied with the existing system, and it did appear as though the system was sufficient for their needs. Paper is a cheap medium, no need for batteries and no problems with sudden "breakdowns". Paper is much easier annotated with data and meta-data, than a PDA or a cell phone, and is not limited by the problems with small and low resolution displays, as well as the lack of efficient input features. Indeed, the limitations in the input features of the mobile devices made it impossible to make quick scribbles and signs, as is the practice with the paper lists, thus rendering the mobile devices useless in the eyes of some of the participating farmers.

Most farmers were, however, very interested in alternative uses for the mobile technology. Mobile technology would be able to provide other solutions, as well as supplement the paper-based tasks, possibly by augmenting them. Especially the *search and identification* (e.g. the ability to show the location of a certain animal by means of a mobile device or identifying an animal by an electronic ear tag) as well as surveillance using cameras were all hot topics among the participating farmers. The overall conclusion of the user study being to use new technology primarily for new tasks, instead of trying to replace the existing paper-based system, which has been in use for more than two decades, and has proven it self highly successful.

4 Conclusion

This paper has described some of the findings of a research project, seeking to discover the potential for mobile computing in the Agricultural sector through a user study consisting of interviews and field studies.

Based on the user study the paper concludes that although mobile technology like PDAs or cell phones may have many potentially interesting uses in the sector, it is not able to replace the most essential tool, the paper-based planning lists. The paper-based media has too many inherent advantages, and most mobile technologies are not yet able to compete with these basic features. However, as mobile device-technology is steadily improving with regards to user interface features and quality, it is likely that these technical limitations may be overcome, in which case the farmers are ready for adopting this new technology.

Mobile technology would, however, be able to provide other solutions, as well as supplement the paper-based tasks, possibly by augmenting them. Especially the search and identification (e.g. the ability to show the location of a certain animal by means of a mobile device or identifying an animal by an electronic ear tag) as well as surveillance using cameras and sensors, were all hot topics among the participating farmers.

References

1. Bardram, J., Kjær T. A. K., and Nielsen, C.: Supporting local mobility in healthcare by application roaming among heterogeneous devices. In *Proceedings of the Fifth International Conference on Human Computer Interaction with Mobile Devices and Services*. Springer Verlag (2003).
2. Beyer, H. & Holtzblatt, K.: Contextual Design, Morgan Kaufmann Publishers (1998).
3. Dourish, P.: Where the Action is – The Foundations of Embodied Interaction. The MIT Press (2001).
4. Kvale, S: Interview - En introduktion til det kvalitative forskningsinterview. Hans Reitzels Forlag (1997).
5. Nielsen, C. and Søndergaard, A.: Designing for mobility: an integration approach supporting multiple technologies. In *Proceedings of the 1st Nordic Conference on Human-Computer Interaction*. Royal Institute of Technology, Stockholm, Sweden (1997).
6. Norman, D. A.: The Invisible Computer: why good products can fail, the personal computer is so complex, and information appliances are the solution. The MIT Press, Cambridge, Massachusetts, USA (1998).
7. Väänänen-Vainio-Mattila, K. & Ruuska, S.: Designing Mobile Phones and Communicators for Consumers need at Nokia. In Information Appliances and Beyond, Chapter 7, Morgan Kaufmann Publishers (2000).
8. Weiser, M: The Computer for the 21st Centrury. In *Scientific American*, September (1991).
9. Weiser, M.: Some Computer Science Issues in Ubiquitous Computing. In: *Communications of the ACM*, July 36, vol. 7 (1993).

Comparison of Mobile and Fixed Use of SmartLibrary

Markus Aittola, Pekka Parhi, Maria Vieruaho, and Timo Ojala

MediaTeam Oulu, University of Oulu
P.O.Box 4500, FIN-90014 University of Oulu, Finland
{markus.aittola, pekka.parhi, timo.ojala}@ee.oulu.fi
mariavie@paju.oulu.fi
http://www.mediateam.oulu.fi

Abstract. Finding books in large libraries can be difficult for novice library users. This paper presents the current status of SmartLibrary, a web-based guidance application helping library customers in this task. We present a comparative evaluation where users tested the service in the main library of University of Oulu with fixed and mobile devices. The evaluation confirmed that SmartLibrary is a useful service for novice library customers; more experienced patrons prefer the traditional shelf classification. The users considered the service easiest to use with a public desktop terminals. However, the possibility of using the guidance in a PDA or a mobile phone in larger libraries was appreciated.

1 Introduction

Finding a target in an unknown environment can be a challenging task. In large libraries, patrons often search for books and other material among the shelves. To aid patrons in this task, we introduced a novel mobile map-based guidance service, SmartLibrary [1]. The user study conducted with SmartLibrary, as well as the survey by Jones et al. [2], showed the need for such a service.

Many libraries provide their customers with OPAC (Online Public Access Catalogues), which is often used as a web service with desktop computers. As an increasing number of customers have their own browser-equipped mobile devices, such as PDAs and mobile phones, it is becoming necessary to provide OPAC also for the mobile Internet. SmartLibrary provided the first OPAC search interface tailored for mobile devices atop of Voyager, a widespread library management system. Compared to desktop computers, however, mobile devices have their limitations (e.g. small screen size, cumbersome keyboard input and limited bandwidth) and strengths (e.g. mobility), which affect the user experience.

In this paper we report SmartLibrary v.2, the re-designed second generation of the SmartLibrary service reported in [1]. Section 2 describes the implementation. Section 3 presents a user evaluation of SmartLibrary v.2 deployed in the main library of University of Oulu. The user evaluation focuses on comparing the user experiences obtained by using the service with a public desktop terminal, a PDA and a mobile phone. Section 4 concludes the paper and discusses future work.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 383–387, 2004.

2 SmartLibrary v.2

The original SmartLibrary prototype was built on SmartWare, an architecture for providing context-aware mobile services [3]. SmartWare provided features such as dynamic positioning of the users and map-based guidance application running in a PDA, among others. The guidance application was integrated with OULA-pda, a web-based OPAC of the Main Library of University of Oulu, customized for mobile devices. The integration was done by adding a hyperlink to the query results provided by the OULA-pda. When the user clicked the link, a separate Java-based guidance application was started.

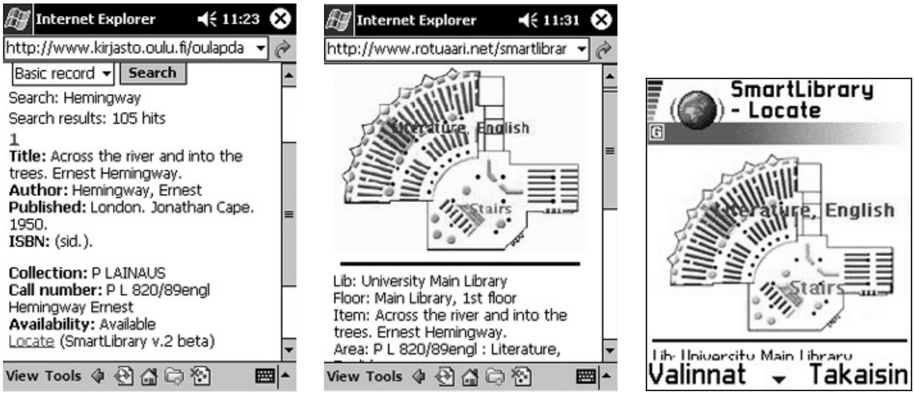
Implementing the first SmartLibrary prototype atop the SmartWare architecture was straightforward, but the prototype had many limitations. The guidance application was slow, and worked only in particular PDA models. Also, the user evaluation of the prototype revealed usability problems. Switching between the web-based OULA-pda and the separate guidance application was considered awkward. The users also had difficulties in orienting themselves on maps, due to poor graphics of the maps.

To overcome the problems of the first prototype, the service has been re-designed according to the web services paradigm. The user interface is provided via the (X)HTML browsers of desktops, PDAs and high-end mobile phones, hence integration with web-based OPACs is seamless. The graphics of the floor plan maps are designed to be simple and clear. Different symbols on the map are color-coded: walls and other fixed structures are drawn with black, book shelves with blue, and tables with yellow. Target areas and their names are superimposed on the map.

To aid users navigating in the library, devices supporting Ekahau's WLAN positioning technology are dynamically positioned within WLAN coverage. Bluetooth cell ID positioning similar to that reported in [4] could also be used. The location of the user is shown as a smiley on the map. In addition to dynamic positioning and fixed structures always drawn on the map, the users can also position themselves relative to pre-defined landmarks shown on the map upon request. The circulation desk and stairs are examples of such optional landmarks.

Fig. 1(a) shows the result of a book search with OULA-pda. By clicking the "Locate" link the user can open the guidance view illustrated in Fig. 1(b). The area of the shelf class ("Literature, English") is drawn with red, and the area of the landmark selected for viewing ("Stairs") is drawn with green. Fig. 1(c) shows the same guidance view in the XHTML browser of a mobile phone.

In large libraries there can be hundreds of shelf-classes and landmarks, whose location may change over time. For the purpose of maintaining this information, SmartLibrary provides the library staff with a web-based Content Provider Interface (CPI). With the CPI the library staff can modify the information of the libraries and their floor plans shown in the guidance UI. The underlying map images of the libraries are also configured using the CPI. The positions of shelf classes and landmarks are drawn on the map as polygons with a drawing tool. Fig. 2 shows a view of the CPI where an administrator is adding a shelf class to the service.



(a) (b) (c)

Fig. 1. (a) A book search with OULA-pda. (b) The guidance view on a PDA. (c) The guidance view on§ a mobile phone.

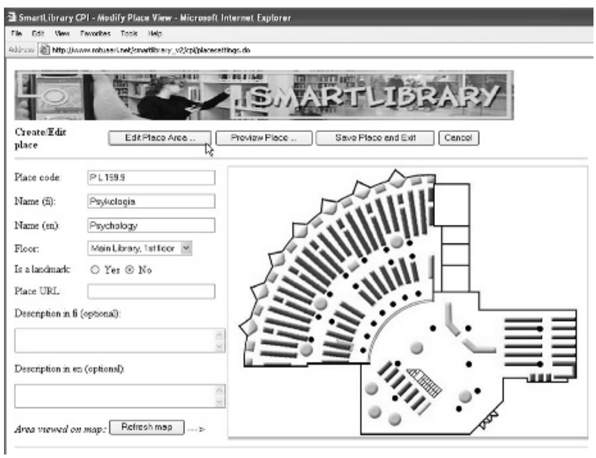


Fig. 2. CPI user interface for modifying a place.

3 User Evaluation

The SmartLibrary v.2 was deployed and evaluated in the Main Library of University of Oulu. 13 voluntary library patrons participated in the evaluation, free of charge. Nine of the test users were professionals or students of library sciences, while four were randomly selected university students. The age of the participants varied between 21 and 50 years, with a median of 26 years. All the users had used OULA before, eight of them on a weekly basis. Three of the participants had used a PDA device earlier. In the test a Compaq iPAQ 3970 PDA with Pocket PC 2002 operating system and a WLAN card was used.

In the initial interview at the beginning of the experiment the users were asked their name, age, and their familiarities with PDA devices, OULA database and the shelf classification system of the main library. The users were then asked to find three books located at the first floor of the library in three different ways: first with the user's own way, second using SmartLibrary with a public desktop terminal and third using SmartLibrary with a PDA. The users were not allowed to ask help from the library staff. The users were asked to think aloud during the search tasks and the whole test was recorded with a microphone attached to the user's collar. The supervisor of the test accompanied the users during the tasks, gave support with technical problems and asked questions about the users' comments. After completing the three tasks the users were again interviewed.

Seven users, out of which six were library scientists, preferred their own familiar way of doing book searches: *"surely own way for it is the most familiar and one has routines using it"*. However, one user working as a library officer would have liked to use the map feature in guiding the customers: *"when it'll be available in the staff's computers, it'll be easy to show the location of the book on the map on the screen to new customers"*. Five users preferred using SmartLibrary with public desktop terminals: *"with desktop, it showed the map anyway and it was more accurate than in the PDA…"* One library scientist preferred using the PDA: *"I liked to work with the pocket PC, could take the map with me"*.

When asked about the usefulness of the service, on four-level scale (1 = useless, 4 = useful) four users gave grade 4, five users gave 3 and four gave 2. On four-level scale (1 = very laborious, 4 = very easy), most users gave grade 4 to SmartLibrary with desktop, and grade 3 to SmartLibrary with PDA. Nine users said that it is useful to see your own location on the map, the more so the larger the library is. The graphics of the map were considered generally clear, but some users found the map too small on the PDA screen. All users found the default structures drawn on the map as useful guidance information, but nobody employed the optional landmarks.

Four of the users tested the service later with a mobile phone, Nokia 6600 featuring Symbian OS, GPRS connection and XHTML browser. Two of the users had earlier experience of using the browser of a mobile phone. Three of the users found it more laborious to use SmartLibrary with a phone than with a PDA, whereas one found them equally easy. The users using the browser of a mobile phone for the first time in the test found text input with the 12-key keyboard of the mobile phone very cumbersome and GPRS connection very slow. After typing three unsuccessful searches to OULA with the mobile phone one user commented: *"if this was my phone, I would throw it to the wall… I would never come to a library just to finger my mobile phone"*.

The evaluation was on purpose partially similar to reported in [1]. In the earlier evaluation, the test users were mostly novice library patrons less familiar with the classification system. In the evaluation at hand, library scientists were in the majority. In both evaluations, most of the novice users preferred using SmartLibrary over the traditional shelf classification, whereas more experienced patrons and library scientists preferred the shelf classification.

4 Conclusions and Future Work

We presented the current status of SmartLibrary, a web-based guidance service for libraries. We compared the user experiences of the service on a public desktop terminal, a PDA and a mobile phone. The user evaluation confirmed the usefulness of the service for novice library users. The test users considered SmartLibrary most useful when used with public desktop terminals. Using the service with mobile devices was regarded less useful, for library customers are more accustomed to using the public terminals, and yet few have a PDA or XHTML-enabled mobile phone of their own. Also, the limited resources of mobile devices, such as smaller screen, cumbersome text input and slower network connection make them less convenient to use. The users considered the mobile version of the service to be more useful in larger libraries.

Currently, SmartLibrary covers the collections, group work rooms and other facilities of the main library of University of Oulu. The library staff is adding other campus libraries into the service with the CPI. SmartLibrary is designed to be used in libraries, but similar guidance service could be useful also for other contexts. We are currently expanding the service towards SmartCampus, a guidance service including numerous lecture halls and other facilities of a large campus.

Acknowledgements. The financial support of the National Technology Agency of Finland and the Academy of Finland is gratefully acknowledged.

References

1. Aittola, M., Ryhänen, T., Ojala, T.: SmartLibrary - Location-Aware Mobile Library Service. Proc. Fifth International Symposium on Human Computer Interaction with Mobile Devices and Services, Udine, Italy (2003) 412-416.
2. Jones, L.W., Rieger, R.H., Treadwell, P., Gay, G.K.: Live from the Stacks: User Feedback on Mobile Computers and Wireless Tools for Library Patrons. Proc. 5th ACM Conference on Digital Libraries, San Antonio, TX (2000) 95-102.
3. Ojala, T., Korhonen, J., Aittola, M., Ollila, M., Koivumäki, T., Tähtinen, J., Karjaluoto, H.: SmartRotuaari - Context-Aware Mobile Multimedia Services. Proc. 2nd International Conference on Mobile and Ubiquitous Multimedia, Norrköping, Sweden (2003) 9-18.
4. Aalto, L., Göthlin, N., Korhonen, J., Ojala. T.:Bluetooth and WAP Push Based Location-Aware Mobile Advertising System. Proc. Second International Conference on Mobile Systems, Applications and Services, Boston, MA (2004) in press.

Automatic Partitioning of Web Pages Using Clustering

Richard Romero and Adam Berger

Nokia Inc.
501 Martindale Street
Pittsburgh PA 15212

Abstract. This paper introduces a method for automatically partitioning richly-formatted electronic documents. An automatic partitioning system has many potential uses, but we focus here on one: dividing web content into fragments small enough to be delivered to and rendered on a mobile phone or PDA. The segmentation algorithm is analyzed from a theoretical and an empirical basis, with a suite of measurements.

1 Introduction

Segmentation is often necessary before transmitting large files to mobile devices. For one, mobile phones often suffer from limited memory and therefore cannot digest large files. Second, gateways (e.g. GGSN products) that mediate mobile data traffic may truncate or refuse to propagate large files. In general, even in the absence of strict file size limits, it is ill-advised to transmit large files over a low-throughput cellular data network, since the recipient may experience an unacceptable latency and/or airtime cost.

There has been limited activity in the field of partitioning richly-formatted documents, but considerably more in the related fields of document summarization and distillation, and in expository text segmentation. The specific contributions of this paper are twofold:

- Introduce a technique, based on clustering, for partitioning richly-formatted content.
- Describe and evaluate a real-time service, based on this technique, which is designed to increase the usability of mobile web browsing.

2 Segmenting via Clustering

This section introduces a tree-based clustering algorithm for document partitioning.

The algorithm takes as input a DOM, the data structure analogue of XML. There exist publicly available parsers that convert XML and HTML files into DOM 2. In addition, there also exist publicly available utilities that convert PDF and Microsoft Office™ file formats into XHTML, an XML-compliant version of HTML. What this

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 388–393, 2004.

means is that the segmentation algorithm described here can easily be adapted to process the most common file formats.

For pedagogic purposes, however, we will assume that the DOM reflects an HTML document: the nodes of the DOM tree correspond to HTML elements like `<p>`, `<b>`, `<table>`, `<td>`, `<it>`, and `<anchor>`, and the leaves correspond to interactive, viewable elements such as text, images, and form widgets.

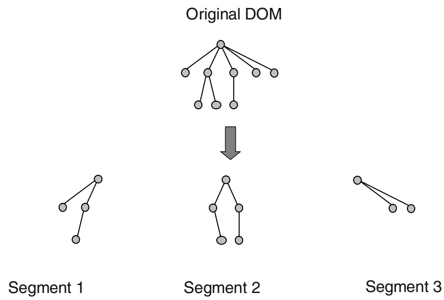


Fig. 1. The segmenting procedure operates on a tree-based document format called a DOM. The input DOM is divided into sub-trees, not necessarily equal in size, but each no larger than a pre-set limit. (DOMs corresponding to real-world documents are, of course, substantially larger.)

One can compute the size of a DOM node recursively; for example, the size of a "b" node is seven bytes (`<b>…</b>`), plus the size of its children. The size of a text node is the number of bytes in the string.

A naïve DOM-segmentation technique is to perform a left-right traversal of the tree's leaves, adjoining leaves to the current segment until adding the next leaf to the current segment would cause the current segment to exceed the specified size threshold. The segment counter is then incremented and a new segment begins. A problem with this approach is that it is insensitive to the inherent structure of the document. For example, it makes no effort to avoid splitting sibling or related nodes: two paragraphs belonging to the same story, for instance, or a heading and the following paragraph. Conversely, it is not positively disposed towards inserting a segment boundary at a natural seam in the page, such as before a long block containing a sequence of hyperlinks.

We now describe a more sophisticated technique, designed to insert seams at "structurally appropriate" locations in the DOM.

2.1 A Clustering Approach

We formulate the DOM segmentation problem as clustering. The algorithm starts by assigning each leaf to its own segment. Adjacent leaves are then aggregated together. Pairs are chosen according to a cost function which encourages merging related nodes (e.g. two adjacent paragraphs) and discourages merging unrelated nodes (e.g. two different frames). The cost function assigns an infinite cost to illegal merges; for

instance, a merge that creates a too-large segment. The algorithm terminates when no finite-cost merges exist.

Algorithm: *Iterative DOM Segmentation*

Input: DOM document D
Cost function *C(x,y)*: a non-negative penalty value for merging DOM segments *x* and *y*

Output: A set of segments, each consisting of a (disjoint) set of contiguous DOM leaf nodes

```
1.     Assign each leaf in D to its own segment.
2.     Compute the cost C(x,y) of merging each adjacent pair
       (x,y) of segments in D.
3.     If C(x,y)=∞ for all x,y then end.
4.     Locate the x=x* and y=y* for which C(x,y) is minimal.
5.     Merge segments x* and y*.
6.     Go to step 2.
```

2.2 Constructing a Cost Function

The cost function *C(x,y)* guides the behavior of the segmentation algorithm. Different assignments for *C(x,y)* result in algorithms with different behaviors and different results.

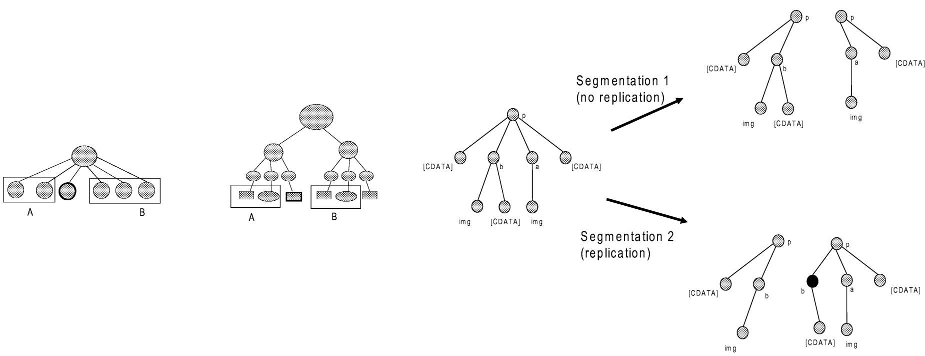


Fig. 2. Considerations in constructing the cost function. **Left:** The unaffiliated leaf can be merged into segment A or segment B. Assuming leaves are equal-sized, merging with A should be lower cost, since that merge balances the sizes of the resulting segments. **Middle:** The unaffiliated leaf node between A or B shares a grandparent with segment A, but only a great-grandparent with B. All else being equal, we expect a lower cost for merging with segment A,. **Right:** Two candidate segmentations of a simple DOM. Notice how in the second segmentation, the black node was replicated, since it is a parent of leaves in both resulting segments. When it comes to node replication, less is better, since replicated nodes increase the total encoding size.

We begin with a simple and intuitive cost function: $C(x,y) = |x| + |y|$. Here $|s|$ is shorthand for "the size of the smallest subtree whose leaves are *s*"; that is, the size of

s plus all the ancestors of s. At a "micro" level, this cost function favors merging smaller segments together; at a "macro" level, the tendency is towards segments of balanced sizes.

Internal DOM nodes may have to be replicated when converting segments into well-formed documents. Often an internal node in the original DOM will appear in two or more segments. The cost function $C(x,y)=|x|+|y|$ overcounts replicated nodes: if node n appears in segment x and y, then it should only be counted once when calculating the size of xy. We therefore need a subtractive term in the cost function, which we denote by $r(x,y)$: the size (in bytes) of the DOM content which appear in both x and y.

Lastly, we observe that two segments related only through a distant ancestor constitute a less compelling merge than two segments related through a parent. To account for this, we append an extra term $d(x,y)$ to the cost function, where $d(x,y)$ is the shortest path in the DOM from x to y.

Many mobile browsers and/or gateways require that the incoming file not exceed a fixed size B. To accommodate this constraint, we require that the cost function assign an infinite cost to the merge of segments x and y when the size of the resulting segment, $|xy|$, exceeds B.

Combining these constraints together, we arrive at a parametric form for the cost function C:

$$C(x,y) = \begin{cases} \alpha(|x|+|y|) - \lambda r(x,y) + \beta d(x,y) & \text{if } |xy| < B \\ \infty & \text{otherwise} \end{cases} \qquad (1)$$

For the experiments reported below, we used α=1, β=100, and λ=1. (β is unitless, but α and λ are in bytes^{-1}.) These values were determined beforehand, by manual optimization on a held-out collection of web content.

Real-world constraints may force a more complex cost function. For example, some mobile web browsers will dispatch HTTP GET requests for only the first n images in a web page, for a browser-specific value of n. After requesting n images, the browsers simply give up, issuing no further GET requests. To avoid this, one could adapt (1) to assign an infinite cost to $C(x,y)$ when the number of `<img>` elements in xy exceeds n.

The algorithm has time requirements of amortized O(n log n) and space requirements of O(n), where n is the number of nodes in the tree. (Space prohibits a full analysis.)

3 Results

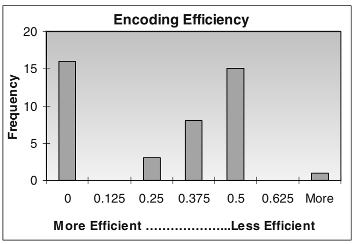
We applied the system to a subset of 41 of the largest web sites from the KeyNote consumer and business web sites[1]. This dataset included URLs as varied as www.dilbert.com, www.intel.com and www.fedex.com/us. For these experiments, B=

---

[1] The full list of 50 is available at www.keynote.com. Nine of these sites were small enough to fit within a single 5KB segment, and were discarded for these experiments.

5000 bytes. The algorithm was executed on a single-processor 2.4GHz x86 machine, on which the segmentation algorithm requires an average of 60ms of CPU time per web page.

The first experiment was designed to measure how close to "optimal" the algorithm is in its segmentation. We define the encoding efficiency as $E=(Nc- N_I)/ N_I$ where N_C is the number of segments generated by the clustering segmentation algorithm for that web page, and N_I is the ideal number of segments: the total size of the web page divided by the pre-set size limit *B*. Note that the *ideal* segmentation doesn't necessarily represent a *valid* segmentation; splitting a web page every *B* bytes is impossible because internal nodes will have to be replicated, pushing the size of the some segments over the limit.



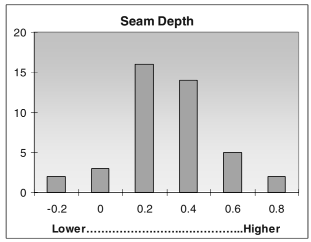


Fig. 3. Two experiments on the KeyNote dataset. **Left:** Encoding efficiency compares the number of generated segments against the theoretical minimum number. **Right:** Relative DOM height of the seam nodes, compared with average depth of a DOM node. A value of zero means the seam appears at exactly the mean depth of an internal DOM node. Generally speaking, higher is better: a seam higher in the DOM separates higher-order structure in the DOM and causes less node replication.

Fig. 4. Real-world example of the partitioning algorithm, this time with B=1400. In this case, the partitioning algorithm generated five segments, outlined in black.

Another quality measure relates to the placement of seams. Generally speaking, seams placed higher in the DOM (closer to the root) are preferable. A seam between two nodes high in the DOM tree is likely to divide two high-level structures: two

tables, for instance, or perhaps a paragraph and a form. A seam between two nodes closer to the leaf level is more likely to split related content: an image and a paragraph, or two text blocks.

Define by $<d>$ the expected depth of a non-leaf node in the DOM, and by d_c the average depth of a computed seam in the DOM. We define the "seam height ratio" as $S = (<d> - d_c)/<d>$. Intuitively, S measures (in relative terms) how much higher the computed seam is in the DOM, compared to the height of a randomly placed seam. The average value of S over the dataset was 0.22, with only four of the pages exhibiting a negative value of S. (In each of these cases, the page had only one seam.)

Also, for illustration, Fig. 4 depicts the behavior of the system on a single web page.

4 Previous Work

Chen *et al* 1 have proposed a system that includes a page-segmentation procedure, which analyzes the document both in DOM *and* in pixel space. The domain of their proposed solution presupposes a certain macro-level structure of a web page (header, footer, left sidebar, right sidebar, body). This approach is not designed to account for network or device-imposed size limits.

Li *et al* 3 propose improving the quality of web search by dividing web pages into cohesive "micro-units," each covering a single topic. The segmentation procedure proposed by the authors involves creating a tag tree (similar to a DOM), and then applying two heuristics to aggregate tree nodes into segments. These two heuristics—merge headings with the following content, and merge adjacent text paragraphs—may be sufficient for creating indexable fragments of the page, but they are too limited for the more general problem of segmentation: the resulting segments will be smaller than desirable, since the algorithm cannot merge segment pairs that don't qualify under these two rules.

References

1. Chen, Y., Ma, W., Zhang H.: Detecting web page structure for adaptive viewing on small form factor devices. *Proceedings of the 12th Annual World Wide Web Conference*. Budapest, Hungary. 2003.
2. HTML Tidy Library Project: http://tidy.sourceforge.net/
3. Li, X., Liu, B., Phang, T., Hu, M.: Using micro information units for Internet search. *Proceedings of the 11th International Conference on Information and Knowledge Management,* McLean, VA. 2003.

Acoustic Features for Profiling Mobile Users of Conversational Interfaces

Dave Toney[1], David Feinberg[2], and Korin Richmond[1]

[1] School of Informatics, University of Edinburgh EH8 9LW.
{dave, korin}@cstr.ed.ac.uk
[2] School of Psychology, University of St. Andrews KY16 9JP.
drf3@st-and.ac.uk

Abstract. Conversational interfaces allow human users to use spoken language to interact with computer-based information services. In this paper we examine the potential for personalizing speech-based human-computer interaction according to the user's gender and age. We describe a system that uses acoustic features of the user's speech to automatically estimate these physical characteristics. We discuss the difficulties of implementing this process in relation to the high level of environmental noise that is typical of mobile human-computer interaction.

1 Introduction

Conversational interfaces allow human users to use spoken language to interact with computer-based information services. Typically, these interfaces are implemented by integrating speech-processing, natural language, and telecommunications systems [1]. An important aim of these systems is to personalize the interaction with respect to the goals, preferences and characteristics of the human user [2].

In this paper we examine the potential for personalizing speech-based human-computer interaction according to the user's physical characteristics. Specifically, we focus on two characteristics of the user: gender and age. In a conversational interface, an estimate of these characteristics can be useful for influencing the style and content of computer-generated utterances. Commercially-orientated services, for instance, can make use of gender differences in consumer behaviour and select their content accordingly [3]. Similarly, elderly users are more susceptible to short-term memory loss. An adaptive interface can use an estimate of a user's age to reduce its navigational complexity [4].

In order to profile a user in this way, a number of acoustic features must be extracted from his/her voice. However, extracting these acoustic features in a mobile context is problematic. The high level of background noise associated with the use of mobile (cellular) phones or in-vehicle devices often restricts the performance of systems based on acoustic feature extraction [5]. In this study we investigate whether an estimation of gender and age is possible within a mobile setting, in spite of the associated background noise.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 394–398, 2004.

In the following sections we propose a set of acoustic features for estimating speaker gender and age. We then outline an implementation of this estimation process and evaluate its performance. We conclude with a summary of our findings and suggest future directions of research.

2 Acoustic Features for User Profiling

What features of the human voice can be used to differentiate one subset of the population from another? More specifically, what acoustic features can help to distinguish between male and female voices, and between younger and older voices? Previous studies [6,7] have identified three acoustically-based features for identifying a person's gender and age: (i) fundamental frequency (F_0), (ii) jitter and (iii) shimmer. The fundamental frequency is the rate at which a person's vocal chords vibrate and is closely related to what is perceived as pitch. Jitter and shimmer, on the other hand, are associated with much more subtle voice qualities. Jitter relates to the periodic variability of the fundamental frequency while shimmer refers to the variation in amplitude of successive pitch periods. Large amounts of jitter and shimmer are often manifested in voices that sound "shaky" or "trembling". In addition to these three features, we make use of a fourth, known as harmonics-to-noise ratio (HNR). HNR is a measure of the amount of noise in a speech signal. Although a high level of noise is expected with mobile communications, a *relative* increase in HNR may indicate older or pathological voices [8].

The measurement of these four voice qualities may be compromised by the noisy environments that are often experienced in mobile communication. To illustrate the effect of noise, Figure 1 contains two spectrograms. A spectrogram shows the variation in energy over time of different vocal frequencies. In this case, the phrase "critical component" was recorded once using a microphone and then again using a mobile phone. The spectrograms of these recordings show the energy variations for frequencies in the range 0–8kHz during a time period of just under one second. It is clear that the second recording contains a great deal of noise.

3 Implementation

We now describe how the acoustic features discussed in the previous section were used to implement an automatic estimator of speaker gender and age. The purpose of the estimator program was to input an example of a user's speech as a `.wav` sound file and output a gender classification and age estimate.

Previous work on age estimation has concentrated on deciding whether a speaker was elderly or not [7]. Acoustic measures such as jitter and shimmer are known to increase appreciably in the elderly. However, in this study we looked at speakers in the age range 21–55. One aim of our study was to investigate whether an estimation of age could be achieved in non-elderly users of mobile (cellular) phones. Previous studies did not investigate mobile phone users [6,7].

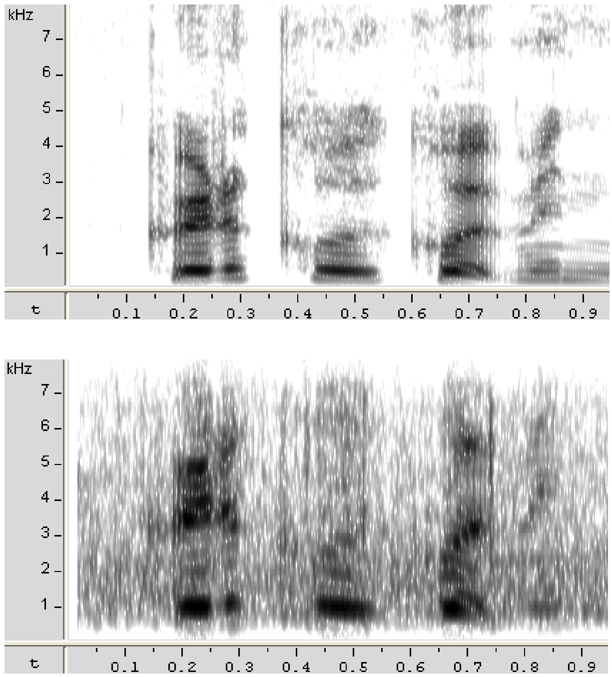


Fig. 1. Spectrograms of the phrase “critical component” as spoken through a microphone (top) and a mobile phone (bottom)

3.1 Tools and Materials

We used two open source tools, Praat [9] and Netlab [10] and one commercial product, Matlab [11]. Praat is an extensive speech analysis and synthesis tool and includes a scripting language for the batch processing of speech files. Netlab is a neural network toolbox that supports a wide range of data analysis techniques. It is used in conjunction with Matlab, a mathematical modelling language [11]. Neural networks provide an effective means of modelling complex relationships between data sources [12]. An important property of neural nets is that their performance is highly resistant to the effect of ‘noisy’ data (i.e. input parameters that contain spurious values).

For a neural network to ‘learn’ how to estimate gender and age, it must be trained using a data set of speakers whose gender and age are known. In our case, we used a subset of the CTIMIT database [13]. This speech database contains single-sentence utterances recorded in a mobile setting. We used only the recordings made by speakers in the age range 21–55, resulting in 3303 recordings spread across 621 speakers.

3.2 Procedure and Evaluation

A Praat script was used to automatically extract estimates of the relevant acoustic features from the recordings. Thirteen measures were extracted for each recording: five values for jitter, six for shimmer, the mean fundamental frequency (F_0) and the mean harmonics-to-noise ratio (HNR). A description of Praat's jitter and shimmer measurements can be found in [14] and [15] respectively. These 13 values formed the input to a multi-layer perceptron neural network which was trained using the scaled conjugate optimization algorithm [16]. All values were normalized into the range [0,1]. The network was trained on 80% of the recordings, validated against a further 10% and evaluated using the remaining 10%. We experimented with a number of network configurations; the best performing network contained 20 hidden layer units.

With respect to gender, the estimator program performed very well. It correctly classified 94.4% of the test cases. In comparison with a simple classifier that always predicted the most frequently occurring value, male (69.1%), the estimator still performed significantly better (two-tailed t-test, $p < 0.01$). Looking at age prediction, the mean error of the test cases was -0.1 years but the standard deviation was 6.86 years. In other words, the neural network failed to learn a useful relationship between the acoustic features and speaker age. We have identified three possible reasons for this result. Firstly, there simply may not be a significant variation in the acoustic features within the age range 21-55. Secondly, the acoustic features that were used may be excessively influenced by background noise. Thirdly, the distribution of speaker age within the CTIMIT database is skewed towards speakers in their 20s and 30s. More training examples of speakers in their 40s and 50s may be required.

4 Summary and Further Work

In this paper we examined a number of acoustic features for profiling mobile users of conversational interfaces. Specifically, we investigated whether a user's gender and age could be estimated in spite of a high level of background noise. We implemented an automatic estimator using acoustic feature extraction and neural network applications. We tested the program using recordings of mobile phone users. The estimator program achieved a very high level of performance with respect to gender but failed to estimate age to a significant level of accuracy.

In the near future, we hope to collect more examples of speakers over the age of 40. This should provide a clearer assessment of the potential for estimating speaker age. We will also investigate other characteristics for profiling users of mobile devices. Potentially valuable user traits include physical size, emotional state, rate of speaking and identification of the user's native language. In the longer term, we intend to integrate these results into a single user profiling module and make it available to developers of conversational interfaces.

References

1. Zue V. and Glass J. Conversational Interfaces: Advances and Challenges. Proceedings of the IEEE **88**(8), 2000. pp 90–169.
2. Zadrozny W., Budzikowska M., Chai J., Kambhatla N., Levesque S. and Nicolov N. Natural Language Dialogue for Personalized Interaction. Communications of the ACM **43**(8), 2000. pp 116–120.
3. Arnold K. and Bianchi C. Relationship Marketing, Gender, and Culture: Implications for Consumer Behaviour. Advances in Consumer Research, **28**, 2001. pp 100–105.
4. Eskenazi M. and Black A. A study on speech over the telephone and aging. In *Proceedings of Eurospeech 2001, Aalborg, Denmark. September 2001.*
5. Gong Y. (1995): Speech Recognition in Noisy Environments: A Survey. Speech Communication **16**(3), 1995. pp 261–291.
6. Minematsu N., Sekiguchi M. and Hirose K. Automatic Estimation of Perceptual Age Using Speaker Modeling Techniques. In *Proceedings of Eurospeech 2003, Geneva, Switzerland. September 2003.*
7. Müller C., Wittig F. and Baus J. Exploiting Speech for Recognizing Elderly Users to Respond to their Special Needs. In *Proceedings of Eurospeech 2003, Geneva, Switzerland. September 2003.*
8. Ferrand C. Harmonics-to-Noise Ratio: An Index of Vocal Aging. Journal of Voice, **16**(4), 2002. pp 480–487.
9. Praat, `http://www.praat.org`.
10. Netlab, `http://www.ncrg.aston.ac.uk/netlab/`.
11. Matlab, `http://www.mathworks.com/products/matlab/`.
12. Bishop C. Neural Networks for Pattern Recognition. Oxford University Press, Oxford, 1995.
13. Brown K. and George E. CTIMIT: A speech corpus for the cellular environment with applications to automatic speech recognition. In *Proceedings of ICASSP 1995, Detroit, US. May 1995.*
14. Measurements of jitter in Praat, `http://www.praat.org/manual/Voice_2__Jitter.html`.
15. Measurements of shimmer in Praat, `http://www.praat.org/manual/Voice_3__Shimmer.html`.
16. Møller M. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks **6**(4). pp 525-533.

Butler: A Universal Speech Interface for Mobile Environments

Botond Pakucs

Centre for Speech Technology (CTT),
KTH, Royal Institute of Technology,
Lindstedtsvägen 24, 100 44 Stockholm, Sweden
botte@speech.kth.se

Abstract. Speech interfaces are about to be integrated in consumer appliances and embedded systems and are expected to be used by mobile users in ubiquitous computing environments. This paper discusses some major usability and HCI related problems that may be introduced by this development. It is argued that a human-centered approach should be employed when designing and developing speech interfaces for mobile environments. Further, the Butler, a generic spoken dialogue system developed according to the human-centered approach is described. The Butler features a dynamic multi-domain approach.

1 Introduction

Recently, the possibility to use speech interfaces in embedded products and consumer appliances in mobile and ubiquitous computing (UC) environments has begun to attract interest. The speech technology industry has already recognized the potentials of the new emerging market. If the market growth of speech interfaces is as large as expected, users will be surrounded by a multitude of speech-controlled services and appliances. However, in mobile environments the usability requirements on speech-based interfaces may increase and new, human computer interaction (HCI) and usability related problems may be introduced.

Some major usability and HCI related issues that should be considered when designing speech-based interfaces for mobile environments are discussed in Section 2 of this paper. In Section 3, it is argued for a human-centered approach and it is suggested that each user should use a single, highly individualized speech interface for accessing a multitude of appliances and services in mobile environments. In Section 4, Butler, a generic spoken dialogue system developed according to the suggested approach is described. Butler features a dynamic multi-domain approach, individualization, user modeling and context awareness.

2 Usability Issues

Speech-based interaction with mobile services differs from accessing speech services through telephones or interacting with desktop computers. Users on the move, and with hands and eyes busy, have greater demands on the HCI.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 399–403, 2004.

Designing and building user-friendly speech-based interfaces with excellent usability properties in their own right might just not be enough. There is a need for a shift in how we think about speech-based interactions in mobile and UC environments. Usability and HCI issues should be considered for whole environments rather than for isolated services and appliances.

2.1 Diverse Speech Technology Solutions

Interface consistency is a central and well-understood concept in the HCI and usability community [1,2]. In mobile environments however, we may expect, in the near future, a multitude of speech interfaces with various complexity, employing a range of different speech technology solutions from simple voice-triggered commands to advanced conversational dialogue systems. A lack of consistency among different speech interfaces may cause usability problems.

When encountering diverse speech interfaces, the same user may be an expert user of some speech interfaces, but still a novice user of other systems. Diverse speech technology solutions may require different interaction strategies from the user and, thus, the use of several different cognitive models. For instance, it will be hard for users to identify the currently available dialogue management strategies, voice commands, and vocabularies. It might even be hard for users to know which services and appliances can be controlled by speech.

2.2 Multiple Concurrent Speech Interfaces

As far as we know, the effects of encountering several concurrent speech interfaces at the same time have never been studied. This situation may actually occur in mobile environments, where several speech-based interfaces are listening for user commands, or even taking initiative pro-actively. Due to miss-recognitions, it is possible that several speech interfaces may be triggered by a single user utterance.

2.3 Increased Usability Requirements

In mobile and dynamically changing UC environments the user's intentions and needs may rapidly change. The user should be able to initiate a new task while waiting for some other specific service to be completed or change the parameters of some previously initiated service. Furthermore, the system itself should be able to interrupt an ongoing dialogue and direct the user's attention to some higher priority events.

For supporting a wide range of domains within one and the same dialogue and for allowing the user to transparently and seamlessly switch between several topic domains and services a *multi-domain approach* [3] is also necessary. The support for these features in current industry solutions is limited.

Consequently, to provide user-friendly speech interfaces in mobile and UC environments and to avoid the introduction of new usability related problems we need means to coordinate and control the various speech interfaces.

3 The Human-Centered Architecture Model

The currently employed speech interface architectures for desktop-based interaction all share an application-centered multi-user system design illustrated in Fig. 1A, where each service or appliance has a separate speech interface [4]. In the fast growing sector of voice portal based telephony services a centralized single entry point can be used for accessing several different services, see Fig. 1B. However, solving the usability problems discussed in Section 2 is not facilitated by these architecture models.

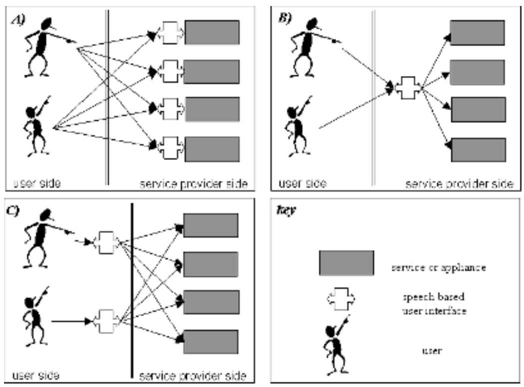


Fig. 1. Speech interface architecture models: *A)* Embedded and application-centered speech interfaces. *B)* Voice portals: - application-centered, centralized speech interfaces. *C)* human-centered and application independent speech interfaces.

The central idea proposed in this paper is the human-centered, application independent architecture for speech interfaces targeting mobile users, see Fig. 1C. Thus, every user is expected to use a *SINGLE*, highly individualized speech interface to access a multitude of services and appliances. It would be preferable if the human-centered speech interface could be integrated into some personal, wearable appliance such as a mobile phone or a PDA. In that case, the speech interface would always be accessible with all user-dependent data activated and ready to use.

Service and application-specific data, including dialogue management capabilities, domain knowledge etc., has to be encoded in *service descriptions* and stored locally at the service provider side. Whenever the user enters a new environment, the available, distributed service descriptions have to be dynamically loaded into the personalized speech interface through some *ad-hoc* and wireless communication solution.

The human-centered, *single user and multiple application* approach to speech interfaces would be an appropriate solution for coordinating and controlling various speech based interfaces. This approach would facilitate the handling of the usability problems discussed in the previous section.

A human-centered approach would facilitate the building of advanced user and domain knowledge models which could provide support for *context awareness* [5]. However, collecting data on the user's behavior, speech patterns etc. is a delicate issue. We believe that a single human-centered interface, because it is controlled by the user, provides better *security and integrity* properties than a multitude of different embedded and distributed systems, which are outside the user's control.

By employing a human-centered solution, it would also be unnecessary for the users to learn and adapt to several different interfaces. The impact of some major challenges for spoken dialogue systems [6] can also be reduced. The speaker variation can be reduced significantly through the possibility to use speaker dependent and speaker adaptive speech recognition. This way the amount of speech recognition errors could be decreased substantially. Addressing challenges such as the variability in channel conditions or background noise could also be facilitated by consistent use of a personalized microphone solutions.

3.1 The SesaME Dialogue Manager

One of the major challenges for the human-centered approach is the dialogue management. SesaME [3] is a generic, task-oriented dialogue manager specially designed for the human-centered approach as well as for mobile environments. Special attention has been given to support adaptive interaction methods and context awareness. In SesaME, a content-based solution [7] is employed for performing the user modeling. This way a simultaneous adaptation to an individual user and to the user's current situation is supported [8].

One of the key-issues in the SesaME architecture is to support a *dynamic multi-domain approach*. The locally available service descriptions, including dialogue descriptions and grammars has to be dynamically loaded and activated on the fly. For handling these requirements, a dynamic plug-and-play functionality of the dialogue management capabilities has been developed [9]. The XML-based service descriptions are distributed through the HTTP protocol however, generic service discovery is also supported.

4 The Butler

Currently, the evaluation of the Butler, a new multi-domain application based on SesaME, is being conducted. The main goal is to evaluate the support for individualization and context awareness, however, speech user interface related problems, such as protecting privacy of the user, disturbance to other people will be also studied. The Butler provides speech-based multi-domain information services through telephones or PDAs. The services provided by Butler can be categorized in three main categories, *public services* such as accessing commuter and subway train timetables, menu information for the nearby restaurants, *accessing personal information* from calendars and *accessing workplace related information*, such as time and location of meetings and seminars.

For identifying the users, telephone number-based or speaker verification is used. The back-end information for all of these services is based on publicly available web-based services. The domain descriptions necessary for the Butler and SesaME are dynamically generated and processed at runtime.

5 Summary and Future Work

Some usability and HCI related problems, which may arise when speech interfaces are integrated in mobile and UC environments have been discussed in this paper. Based on these issues, a novel human-centered approach is proposed for speech interfaces in mobile environments. Further, SesaME, a generic multi-domain dialogue manager, built according to the human-centered approach, has been described. The SesaME dialogue manager is employed in the framework of the Butler demonstrator. By employing a dynamic multi-domain approach, the Butler acts as an individualized universal speech interface.

The suggested approach creates novel possibilities for supporting personalization, context awareness and user modeling in dialogue management. These features will be studied in an upcoming long-term user-study.

Acknowledgments. This research was carried out at the CTT, a competence center at KTH, supported by VINNOVA (The Swedish Agency for Innovation Systems), KTH and participating Swedish companies and organizations. This work was also sponsored by the European Union's IST Programme under contract IST-2000-29452, DUMAS (Dynamic Universal Mobility for Adaptive Speech Interfaces).

References

1. Grudin, J.: The case against user interface consistency. Communications of the ACM **32** (1989)
2. Nielsen, J.: 5.4 Consistency. In: Usability Engineering. Morgan Kaufmann, Inc., San Francisco, CA, USA (1994)
3. Pakucs, B.: Towards dynamic multi-domain dialogue processing. In: Proceedings of the Eurospeech'03. Volume 1., Geneva, Switzerland (2003) 741–744
4. Larson, J.A.: Speech-enabled appliances. Speech Technology Magazine (2000) http://www.speechtek.com/.
5. Dey, A.K.: Understanding and using context. Personal and Ubiquituos Computing **5** (2001) Special issue on Situated Interaction and Ubiquitous Computing.
6. Glass, J.R.: Challenges for spoken dialogue systems. In: Proceedings of 1999 IEEE ASRU Workshop, Keystone, CO, USA (1999)
7. Zukerman, I., Albrecht, D.W.: Predictive statistical models for user modeling. User Modeling and User-Adapted Interaction **11** (2001) 5–18
8. Pakucs, B.: SesaME: A Framework for Personalised and Adaptive Speech Interfaces. In: Proceedings of EACL-03 Workshop on Dialogue Systems: Interaction, Adaptation and Styles of Management, Budapest, Hungary (2003) 95–102
9. Pakucs, B.: VoiceXML-based dynamic plug and play dialogue management for mobile environments. In: Proceedings of ISCA T&R Workshop on Multi-Modal Dialogue in Mobile Environments, Kloster Irsee, Germany (2002)

Learnability and Automatic Clustering of Images in Mobile Devices

Aino Ahtinen and Malin Andersson

Nokia Research Center
P.O. Box 407, FIN-00045 Nokia Group, Finland
{aino.ahtinen, malin.andersson}@nokia.com

Abstract. Today's imaging phones create new challenges of managing large amount of images. The user has to be able to browse, find and organize media in an effortless way on a small display and with limited navigation possibilities. We present a mobile application, which clusters images automatically based on date and location. We conducted a user study with the application in order to find out what users think of automatic clustering and how they learn to use it. The results revealed that our application with automatic clustering of images helps users to manage a large amount of images.

1 Introduction

There are several applications for managing and viewing digital images on PC, e.g., Adobe Photo Album [1], Canon Zoombrowser [3], and Photo Mesa 2.01 [2]. Adobe Photo Album displays the images in a chronological order using time stamps when the photos were taken and it offers different possibilities to zoom in and out. Canon Zoombrowser allows the user to organize photos into different folders and displays the content of folders by using thumbnails. The user can use a zoom function to enlarge desired content of one folder or one single photo. Photo Mesa uses also a zoomable user interface to optimize the screen when displaying photo collections but it requires the users to organize the photos.

Recent studies highlight the importance of effective tools for photo management [4, 5,7]. Automatic clustering of digital photos is one enabler for efficient management [4]. It can be a powerful tool for organization, but clustering cannot provide high user satisfaction alone. A good user interface that supports the concept of automatic clustering is therefore very important [8].

Graham et al. [6] stress the importance of displaying time in the user interface in order to support management of images. They developed an application that automatically clusters images, which are taken in certain time period, as events. Their study pointed out that browsing among clustered images was easy due to events.

An essential characteristic of imaging phones is mobility and, therefore, one of its primary attributes is location information. The location information of places where images are taken offers new possibilities to cluster images. Our approach builds on

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 404–408, 2004.

the concept of clustering images automatically based on time and location to support browsing. The concept was tested with the prototype application. How the test participants managed to use and understand the clustering based on these two attributes will be discussed further in this paper.

2 Automatic Clustering by Time and Location

For the user study we had a prototype application in the mobile phone. The prototype clusters images automatically based on date and location. Date and location are metadata attributes, which are included in the file when the image is created. Images taken in the same location and time, from 00.00 am to 11.59 pm, are grouped together unlike the approach that Graham et al. [6] use in their photo browser. Their photo browser automatically organizes photographs by events. It realizes that the sequences of photos taken closer in time are one event [6].

The location information is based on GSM network cells. The cell is the area covered by one base station in GSM network. Cell-based positioning may have problems with accuracy. For example, positioning in urban areas is more accurate due to smaller cell sizes, whereas, positioning in countryside may be inaccurate because of large cell sizes. Furthermore, in urban areas cells overlap.

Our application names images with a default name based on location (e.g., “Location 1”). Obviously, the user can rename these locations and after renaming, all images, both existing and future, taken in this location will be grouped and named after it. The personalized location names distinguish the different clusters from each other and support the user in remembering the capturing location. Clusters with different cell-id:s can be set to same location name, which gives more flexibility for the user to determine and personalize how the location information can be used.

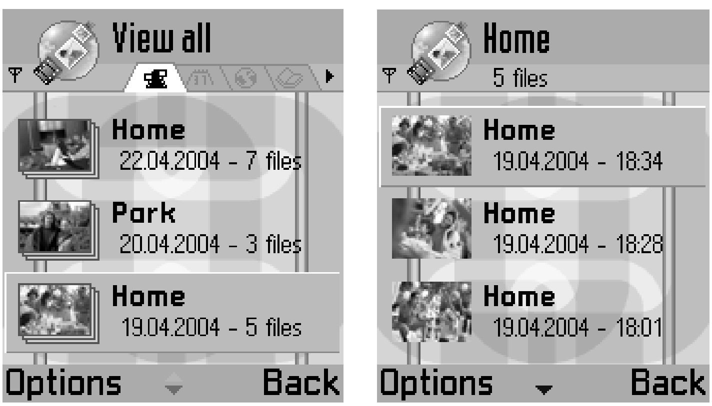


Fig. 1. A list of three clusters (left) and inside one cluster (right)

Figure 1 presents two different views of our prototype. On the left is a cluster view, which provides searching tips for the user: thumbnail of the last image taken in that

cluster, location name, creation date and number of files inside the cluster. By opening the selected group, the user can access the images taken in that location at the same day. On the right is an opened cluster with images in that cluster.

3 The User Study of Automatic Clustering

The focus of our user study was to find out what users think of and how they learn to use the automatic clustering of images based on time and location in a longer-term usage.

3.1 Method

We conducted our study with 12 participants (6 male and 6 female). Participants were experienced users of mobile phone and camera application. Their average age was 28 years. All participants were Finnish, and they came from various backgrounds.

The study was organized as follows: First we carried out an initial usability test and semi-structured interview; then the subjects had a free usage period of two weeks and finally we arranged another usability test and interview. The initial and final usability tests included 14 tasks that were mainly about searching images. The tasks were comparable in the initial and final usability tests. The tests were carried out with the application running in the Nokia 6600 mobile phone. The usability test phone included a set of about 250 images, and thus, all participants conducted the tasks with the same content. However, the phones that users used during the usage period of two weeks did not include any beforehand created content.

We studied learnability of the application by measuring task completion times and usability problems before and after the 2-week-usage. By interviewing participants we got more detailed view of how users arranged image-files (e.g., on PC), and how they perceived and used automatic clustering of images in the mobile device. Interviews were analyzed qualitatively, i.e., by using content analysis.

3.2 Results

In the typical image gallery application of the mobile phone, images are displayed as a list (see, e.g., Media Gallery of Nokia 6600 phone), and there is a possibility to organize them into folders. According to the initial interviews, users do not usually create folders for images on the mobile phone image gallery. Consequently, images are left as they appear – in one and, usually, long list. The users find this problematic, because it makes searching of images difficult. Therefore, the users would like to have their images grouped to subfolders, preferably automatically.

When the number of images in the mobile phone increases high enough, users usually transfer them to the computer and organize them into folders. We studied how users manage a large amount of images on the PC and found out that the most used way to organize images is to group them based on events, e.g., a holiday trip or a party.

The user interviews carried out after 2-week-usage period revealed four main advantages in the automatic clustering of images based on time and location compared to the normal list view of the typical image gallery of the mobile phone. Firstly, there is no need to use any effort for clustering images, since the clustering is automatic. Secondly, automatic clustering shortens the list, which is essential on a small screen. Next, transferring images from phone to computer is easier as the user is able to transfer groups. Finally, clustering makes searching of images faster, because user does not have to scroll all images one by one. Clustering helps a lot in searching especially when the user has named at least the most important locations.

However, we found out also some problematic areas. As images form groups, it makes 'photo album style' browsing more difficult, because user has to open and close groups to view images one by one. In addition, the best way to use clustering requires still some effort from the user because s/he has to rename locations. If the user does not name locations, one aid for recognizing group is missing due to the vague default names. Recognizing a group depends also on the thumbnail of the group and if it is unrepresentative, it may mislead the user. In some situations, the cell areas do not match with how users differentiate locations. Inevitably, there is a mismatch between technology driven network cell positioning and the way that human beings define their environment in places. However, the users commented that the automatic clustering of images is so beneficial that they would like to use it despite of these problems.

Understanding of the clustering logic usually required learning by doing. Some users did not understand on which parameters the clustering was based in the beginning. However, all users understood how the groups are formed after the two-week usage period. Based on the usability test task completion times, users performed on average 29% faster in the final test compared to the initial test. By observing the tests, no serious usability problems related to the automatic clustering were found in the UI. Also, users rated test tasks, on average, as easy.

As mentioned earlier, users often arrange images based on events on PC. According to our interviews after the usage period, it seems that users would like to have images arranged by events also in the mobile phone. According to the participants, one event can consist of images taken in several locations.

4 Conclusions

We have presented an application that manages a large amount of images by clustering them automatically based on time and location. Our user study revealed that automatic clustering of images by location and time is beneficial for the users when storing a huge amount of images in a mobile phone. The greatest advantage in the automatic clustering is that it makes searching of images easier when compared to the normal list view of images. The logic of clustering was easy to understand for the users after using the application for a while. However, further study is needed to find out, whether clustering by time and location is the best way to group images. It seems that users would like to group images more by broader events than single locations,

which was also suggested by Graham et al. [6]. This could support the concept of events constructed by several together-belonging locations. From the user's perspective, there could for example be an event called "Holiday trip" that could include subfolders, i.e., image-groups taken in several locations during the trip.
Most users wanted to have also a traditional list view in those cases when they just wanted to browse images, not search a certain image. By offering different clustering parameters and a possibility to turn it totally off, users can choose how to use automatic clustering to better support their needs. Clusters simplify image organization due its character – grouped images with the same metadata. Marking and selecting single images, one by one, is both time consuming and frustrating in the long run. Clusters can be used, e.g., when creating collections, sending, sharing, and publishing images. In the future, when users can use other valuable metadata and choose among different clustering principles, the automatic clustering will be indispensable.
Imaging phones should make use of the fact that images are taken in different locations and during different events. Locations mirror places of capture, which can be very useful information when organizing images. Automatic clustering by location and date emphasizes the essential character of the imaging phone - its mobility.

References

1. Adobe Photo Album, http://www.adobe.com/products/photoshopalbum/main.html
2. Bederson, B.B., Quantum Treemaps and Bubblemaps for a Zoomable Image Browser, ACM Conference on User Interface and Software Technology (UIST 2001) as PhotoMesa: A Zoomable Image Browser using Quantum Treemaps and Bubblemaps, pp. 71-80. May 2001
3. Canon Zoombrowser, http://www.canon.com/technology/detail/digi_camera/original_soft/index.html
4. Cooper, M., Foote, J., Girgensohn, A., Wilcox, L.: Temporal event clustering for digital photo collections, Proceedings of the eleventh ACM international conference on Multimedia, November 02-08, 2003, Berkeley, CA, USA
5. Frohlich, D., Kuchinsky, A., Pering, C., Don, A., and Ariss, S. Requirements of Photoware. Proceedings of the 2002 ACM conference on Computer Supported Cooperative Work (CSCW), New Orleans, Louisiana, USA, ACM Press, 2002, pp.166 – 175
6. Graham, A., Garcia-Molina, H., Paepcke, A., Winograd, T.: Time as essence for photo browsing through personal digital libraries, Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries, July 14-18, 2002, Portland, Oregon, USA
7. Platt, J. C.: AutoAlbum: Clustering Digital Photographs using Probabilistic Model Merging. Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'00), p.96, June 16-16, 2000
8. Platt, J. C., Czerwinski, M., Field, B. A.: PhotoTOC: Automatic Clustering for Browsing Personal Photographs, Microsoft Research, USA. February 2002

UbiquiTO: A Multi-device Adaptive Guide

Ilaria Amendola[1], Federica Cena[1,2], Luca Console[2], Andrea Crevola[1], Cristina Gena[2], Anna Goy[2], Sonia Modeo[1,2], Monica Perrero[1], Ilaria Torre[2], and Andrea Toso[1]

[1] CSP innovazione nelle ICT s.c. a r.l., Torino, Italy
{amendola, cena, crevola, modeo, perrero, toso}@csp.it

[2] Dip. Di Informatica, Università di Torino, Torino, Italy
{lconsole, cgena, goy, ilatorre}@di.unito.it

Abstract. This paper describes UbiquiTO, an adaptive tourist guide, conceived as a "journey companion" for mobile users in Turin, aimed, for the current prototype, at supporting mobile workers helping them to organize their late afternoon and evening in town. The paper is intended to emphasize the most relevant feature of the system, that is the integration of different adaptation strategies in order to allow high flexibility in terms of device used, localization technology, user preferences and context conditions.

1 Introduction

The convergence of pervasive computers and communication networks offers new opportunities and challenges for systems designers and the exponential diffusion of devices such as PDAs and smart phones drives in the same direction, but the opportunity to use and profit from digital services depends on the possibility of adaptation to the mobile context, including input/output modalities, goals and location of the user, and so on. In particular, we think that the last feature represents a real add-on to common digital services, enabling the so-called *location based services*. Moreover, in order to make mobile services really useful and profitable, the characteristics and preferences of the users should be taken into account.

In this paper we describe UbiquiTO, an agent-based expert tourist guide for mobile users (the first prototype focuses on mobile workers), filtering the information and delivering it in the most appropriate way, depending on different factors: (a) Tourist services are provided according to the *location* of the user. (b) The user interface adapts to different types of *devices*; in the current prototype, we focused on PC and PDA. (c) The system exploits *user profiles*, including her interests, preferences and her previous visits to Turin, in order to provide personalized suggestions. (d) The interaction is adapted taking into account a set of *context* parameters such as the time of the day, the fact that the user is moving, and so on. Moreover, services are provided in two different ways: as consequence of explicit request from the user, who asks for a specific support (e.g., to find a hotel or a restaurant, or to get information about events or places of interests); by proactive activation, when the system itself, in

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 409–414, 2004.

specific situations, depending on the adaptation strategies mentioned above, autonomously provides the user with tourist advices. In the following, Section 2 provides an overview of the system architecture; Section 3 presents some details about the different forms of adaptation; Section 4 concludes the paper and compares UbiquiTO to some related works.

2 UbiquiTO Architecture

UbiquiTO architecture includes four main agents: the *Recommender* exploits the personalization rules to suggest items tailored to the user preferences and location; the *Presentation Adapter* exploits the adaptation rules to adapt the presentation (e.g. descriptions) to the user preferences, the device characteristics, and the context; the *Interaction Manager* handles the dialog with the user: each dialog step corresponds to the generation of a XML object representing the personalized content of the page to be displayed; the *UI Generator* handles the application of XSL stylesheets that transform the XML object into the (X)HTML pages representing the User Interface (UI), taking into account the different characteristics of the devices and the context features.

The architecture also includes four specialized modules, which handle, respectively, the user profile, the model of the device, the information about the location and a model of the context (environment conditions). Each specialized module exploits specific features stored in two main databases: the *users DB* and the *places DB*.

3 Adaptation Strategies

As several studies suggested (e.g., [3]), adaptation techniques can be effectively exploited to handle the interaction in mobile user interfaces. In the current prototype, two user interfaces have been designed: one for PC and one for PDA (see Figure 1). The system can generate an adaptive version (AV) and a non-adaptive version (NAV) of both user interfaces.

Adaptation of Content. In the AV, in order to suggest places to visit, restaurants, accommodations and so on, the Recommender assigns a score to each item and orders them. The computation of this score takes into account: (a) the user's interest in the category the item belongs to; (b) the proximity of the item to the user position, in case the user exploits a mobile device. In the NVA, the items are ranked only according to their popularity. Moreover, in the AV, when the system is asked to provide an item description, it adds a list of suggestions tailored to the user profile and the user location (see right-hand side of PC and PDA user interfaces in Figure 1).

Fig. 1. UbiquiTO user interfaces: PC, on the left and PDA, on the right.

Adaptation of the User Interface. The Presentation Adapter personalizes: (a) the amount of information displayed, according to the screen size of the user's device; (b) the font size and background colour, according to context conditions (e.g., time of the day, movement or not) and user features (age, possible vision impairments, etc.).

Localization Strategies. In the UbiquiTO project we considered four methods to localize the user and in the current prototype we implemented the first two ones:

1. *User-driven localization.* If the user is not equipped with a device supporting automatic positioning, like GPS, she has to provide the system with the coordinates of the point she is closest to. The UbiquiTO UI offers two different ways to specify the user's location: she can select a point from a list of items, or she can click on a sensitive map. The interaction with the map involves two steps, at different levels of detail: in the first step, the system shows the user a map representing the whole area (the center of Turin) and some of the most important points of interest in town (monuments, churches, etc.); in the second step, a more detailed map, representing a zoom of the previously selected area, is shown, containing a larger number of points of interest. In both cases, the user clicks on the point she is closest to, the system retrieves its coordinates from the places DB and computes the coordinates of the user position as approximately corresponding to those of the selected point. Notice that this form of localization may also be exploited by users to ask for information non current-location-dependent.
2. *Wireless LAN.* If the user mobile terminal is equipped with a WiFi receiver and enters in wireless modality, her position can be computed on the basis of the signals received from the different access points within the area. This positioning method is rather precise, but it can be exploited only within areas covered by wireless LAN. We have planned to test this kind of localization method inside the Environment Park (in Turin), in cooperation with INLAB (see [8]).

3. *GPS*. In this case the tourist's device contains a GPS receiver that enables the system to calculate the user position with great accuracy. Unfortunately, the GPS technology is affected by some limitations: its low diffusion; the slowness of the survey; the bad performance indoor.
4. *Network-based positioning*. Mobile networks operators can compute the user position quite precisely, on the basis of the distance of the user from the closest cell, even though the accuracy depends on the cells size (the smaller are the cells, the more accurate is the localization). Moreover, the main problem of this positioning method is represented by privacy restrictions, which in some countries, including Italy, prohibit telecom companies from releasing this kind of information.

Independently of the elicitation method used, the user position is represented by a couple of coordinates. Given this information, the system retrieves, from the places DB, the coordinates of places to be recommended and calculates the distance between user position and every single place, in order to suggest the closest ones.

4 Conclusions and Related Works

In this paper we have presented UbiquiTO, an expert tourist guide for mobile users that adapts the content provided and the interaction to the user interest and physical location, as well as to the devices and context conditions. Mobile guides typically merge approaches from different fields. UbiquiTO borrows techniques from User Modelling and combines them with wireless technologies. The integration of different adaptation strategies is probably the most relevant aspect of UbiquiTO, since it supports a high flexibility, by adapting to heterogeneous factors. Several works in different areas are significantly related to the project. Even restricting the analysis to mobile guides, the number of systems developed since the first prototype, Cyberguide [1], is high (see Guide [4], Lol@ [11], Crumpet [10], Real [2], SmartKom [12], Deep Map [7], as a sample of the main ones). Therefore, we will take into consideration just those ones that are most comparable with the main features of our system.

As said, the most important characteristic of UbiquiTO is its flexibility, due to the integration of device and location adaptation with adaptation to the user features. This combination enables the tourist to use her own mobile terminal, not necessarily equipped with specific positioning devices or client-side applications, and to benefit from the advantages of the adaptation. From this point of view, Crumpet [10] is probably the most relevant related project. It personalizes services to the user's current location, interests, and history of interaction, on PDA and smart phones; it also adapts the presentation to changing technical environments and exploits GPS or other operator based technologies (e.g. GSM, UMTS) to localize the user. Guide [4] is another relevant project. It is a tourist guide for the city of Lancaster, adaptive to the location of the user, her walking speed, the places already visited, the time of the day and the language and interests of the user. The main lack of the system concerns the client device: Guide services may be accessed only using an ad hoc terminal rented at

the Lancaster tourist office. Lol@ [11], a guide for the city of Vienna, is adaptive toward the device, but not toward the user features. It uses GPS as positioning technology and exploits a GIS system for the generation of the map. As in UbiquiTO, the layout is adapted using XSL stylesheets. In spite of its complex infrastructure, the lack of adaptivity toward the user interest, preferences and goals may compromise, as emphasized by [3], the possibility to enjoy the services in mobile environments. In term of flexibility, one last feature we would like to briefly discuss is the modality to elicit the user position. A good half of the outdoor systems uses GPS, while just a few allow the estimation of the position by interacting with the user. Guide, Lol@, Real and Deep Map are some examples of integration of alternative methods to elicit the user position [6]. As seen, one of the goals of UbiquiTO is to go in the same direction: the current prototype offers two modalities: layered maps and WiFi technology; in the next versions, we will experiment GPS modality and possibly Network-based solutions. A relevant aspect of the map is that it is built upon the concept of *points of proximity* which shares the principles of landmarks, already experimented in several contexts (see [9]). The main idea is that they represent relevant points in the user mental map and help the user to construct a mental representations of unfamiliar environments.

The first version of UbiquiTO is currently ready to be tested. We have planned a layered evaluation of the system, aimed at testing both the adaptation of the content and the adaptation of the user interface. Many aspects will be improved in future versions. Different mobile devices will be taken into account (e.g., smart phones, on-board equipments) and a larger set of services will be included. Moreover, in order to automatically update the user profiles, learning mechanisms will be studied and implemented. Finally, localization mechanism exploiting GPS will be experimented.

References

1. Abowd, D. A., Atkeson, C. G., Hong, J., Long, S., Pinkerton, M.: Cyberguide: A Mobile Context-Aware Tour Guide. Wireless Networks 3(5) (1996) 421-433
2. Baus, J., Kray, C., Kruger, A.: Visualization of route descriptions in a resource-adaptive navigation aid. Cognitive Processing 2(2-3) (2001) 323-345
3. Billsus, D., Brunk, C., Evans, C., Gladish., B., Pazzani, M.: Adaptive Interfaces for Ubiquitous Web Access. Comm. of the ACM 45(5) (2003) 34-38
4. Cheverest, K., Davies, N., Mitchel,l K., Smyth, P.: Providing tailored context-aware information to city visitors. In Adaptive Hypermedia and Adaptive Web-Based Systems (AH 2000). LNCS Vol. 1892, Springer-Verlag, Berlin Heidelberg New York (2000) 73-85
6. Kray, C., Baus, J.: A survey of mobile guides. Workshop on HCI in mobile guides, 5th Int. Symposium on HCI with Mobile Devices and Services. Udine, Italy (2003)
7. Malaka, R. and Zipf, A.: DEEP MAP - Challenging IT research in the framework of a tourist information system. Information and Communication Technologies in Tourism. Proc. of ENTER 2000, Springer Computer Science, Wien New York (2000) 15-27
8. Marchioro, C.E., Ghittino, A., Ferri, D., Borri, R., Dovis, F.: Wireless Campus project. TERENA Networking Conference 2003. Zagreb, Croatia (2003)

9. Michon, P. E., Denis, M.: When and why are visual landmarks used in giving directions. In: Montello, D.R. (ed.): Spatial Information Theory. Lecture Notes in Computer Science, Vol. 2205, Springer-Verlag, Berlin Heidelberg New York (2001) 292-305
10. Poslad, S., Laamanen, H., Malaka, R., Nick, A., Buckle, P., Zipf, A.: CRUMPET: Creation of User-Friendly Mobile Services Personalised for Tourism. 3G Mobile Communication Technologies. London, UK (2001) 26-29
11. Pospischil, G., Umlauft, M., Michlmayr, E.: Designing LoL@, a Mobile Tourist Guide for UMTS. Information Technology & Tourism 5(3), Cognizant Communication Corporation (2003) 151-164
12. Wahlster W., Reithinger N., Blocher A.: SmartKom: Multimodal Communication with a Life-Like Character. 7th European Conference of Speech Communication and Technology (Eurospeech 2001), Vol. 3, Aalborg, Denmark (2001) 1547-1550

Visualizing the Thematic Update Status of Web and WAP Sites on Mobile Phones*

Luca Chittaro

HCI Lab, Dept. of Math and Computer Science,
University of Udine
via delle Scienze, 206, 33100 Udine, ITALY
chittaro@dimi.uniud.it
http://hcilab.uniud.it

Abstract. The primary goal of people accessing the Web from mobile phones is to find specific pieces of information (PoI, hereinafter), not to surf. Well-designed sites for mobile users help them by minimizing the path needed to reach the desired PoI. We propose a further improvement, based on visualizing thematic update status (i.e., how many PoI have been added in each category and when). This can prevent unfruitful navigation of the site and also allow users to compare different sites to choose which one better suits their needs.

1 Introduction and Motivations

All recently proposed guidelines [3,5,6,7] for designing Web and WAP sites aimed at mobile users agree that minimization of navigation time to reach information is crucial. Moreover, they stress that mobile users have different goals, tasks, and constraints than users sitting in front of a computer, and their primary goal is to find specific pieces of information (PoI, hereinafter), not to browse the Web [5,6]. Well-designed sites for mobile users focus thus on rationally organizing the different PoI into meaningful categories and minimizing the length of the navigation path needed to reach any category and PoI. In this paper, we concentrate on a further improvement that aims at saving additional time by making the user aware of the thematic update status (i.e., how many PoI have been added in each category and when) of the site.

As a representative case study of mobile sites, we analyzed the most popular international news sites for mobile phone users [1,2,4,8,10]. Their design solutions and navigation path to access PoI are identical. Fig. 1 illustrates an example of a user accessing an information that is important for investments. The user selects (Fig. 1a) the proper category (if there are subcategories, additional selections are needed); she is presented with a list of titles for the selected category (Fig. 1b); she scrolls until she finds a specific title and chooses it; date and text of the chosen PoI appear in a new page (Fig. 1c). The only differences among the sites concern date information (two sites [4,8] do not show dates of the news) and advertising (based on light graphics) that lasts a couple of seconds before the selected news is shown (only in [2]).

* This work has been partially supported by the MIUR COFIN 2003 program.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 415–420, 2004.

(a) **(b)** **(c)**

Fig. 1. Accessing a PoI on mobile news sites from a mobile phone.

Although this design is well-thought, there is still room for improvement, especially to meet the needs of regular, frequent users of sites (this class of users is considerable for sites that are continuously updated such as news, finance, sports, weather sites,…). A problem that comes up when using these sites over a period of time is insufficient awareness about thematic update status. As a result, when the user visits the site, she has to check the desired categories and title listings for new PoI, even if she has already seen them in previous visits. As a practical scenario, consider a manager on a trip to a meeting in a distant city, who needs to periodically visit different sites to: (i) read the latest business news, (ii) look for possible changes affecting her flights or those of the people she has to meet, (iii) be informed about the latest weather forecasts to decide if she wants to book a tour to the park close to the destination. Knowing *if* categories of interest have been updated after the last visit would prevent useless navigation. More generally, users should be also made aware of *when* categories have been updated. For example, if a user is looking for the results of on-going sports competitions or shares in the stock-market or the progress of a military crisis, knowing that updates have been made in the last minutes makes them more relevant than those made hours ago. This is thus useful also to users who are not frequent visitors of sites, and can help in choosing what is the best site to visit for the purpose.

A traditional solution to the considered problem could be based on alerting services, but requires users to register to the service and choose which updates could possibly be interesting and should trigger alerts. The user has to repeat the process on all the sites she visits. Alerting can have the undesired effect that users interested in many topics and/or sites might find their mobiles flooded by alerts (and possibly additional unwanted messages). Moreover, it would be very difficult to get a picture of the status of interesting sites by trying to relate a list of separate alerts that concern only some changes (limited screen space makes this task even harder, forcing users to jump around through multiple screens). A more sophisticated solution would maintain a database that tracks what each user has read. Although this could allow the user to get a detailed account of the unread updates of interest, it could be inconvenient both for users (not everyone would be willing to register and login to all the sites she visits) and sites (not every site would be glad to maintain large databases of individual usage information and force users to register and login to get the new functionality).

For the above reasons, we studied a solution that is not based on alerting, is available to any site visitor without registering and aims at quickly communicating a clear picture of thematic update status. Users should be able to stop at the first of the 3 phases in Fig. 1, and proceed to the following phases only if the status information in the first phase motivates them to do so. Besides time savings, instant awareness of thematic update status allows the user to compare different sites to choose which one better suits her needs (e.g., the one that devotes more attention to a given category of information, the one with more recent updates to a given category,…). The solution we describe can thus be useful also for users who are not frequent visitors of sites.

2 Visualizing Thematic Update Status

We employ simple but informative graphics to present thematic update status at-a-glance, in the same page that lists categories. Graphics have to be small and simple [5,6], so that they can be drawn on a limited display and quickly downloaded. We also: (i) base our visualizations on well-known graphic elements (such as bar and pie charts) that are familiar to users, avoiding to extend them with unnecessary graphics that harm their readability (see [9] for a discussion), and (ii) use a limited number of colors that are easy-to-distinguish (also on those color phones that do not render well).

2.1 Representing Temporal Information

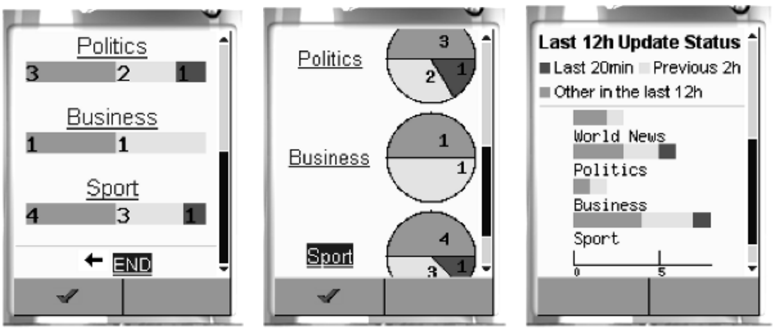
Communicating thematic update status requires to refer to time, choosing (i) the most appropriate time intervals, (ii) the most appropriate words to name the intervals. Time intervals can be either *disjoint* (e.g., the last 5 minutes and the 25 minutes that preceded them are disjoint intervals) or *overlapping* (e.g., the last 5 minutes and the last 30 minutes are overlapping intervals). The intervals extent is also important: how many intervals, and how wide, are both useful and easy-to-understand for users? To define temporal aspects, we interviewed 30 subjects who use mobiles and computers, asking them to imagine a fictitious site that provides thematic update status in the format they would find more useful. Most interviewed subjects organized information in 3 intervals of time. The average periods of interest were around the last 20 minutes, around 2 hours and around 12 hours. There was less consensus about the type of intervals: although more than half of users reasoned in terms of overlapping intervals, a considerable part of them referred to disjoint intervals. We thus designed some visualizations based on overlapping and some on disjoint intervals. With disjoint intervals, we divide the chosen 12-hours timespan into 3 intervals called *Last 20min* (corresponding to interval [-20,0] in minutes, where 0 is current time), *Previous 2h* ([-141,-21]), *Other in the last 12h* ([-720,-142]). With overlapping intervals, the 3 intervals are *Last 20min* ([-20,0]), *Last 2h* ([-120,0]), *Last 12h* ([-720,0]).

The interviews also explored color coding for the 3 intervals. The preferred approach was a traffic light scheme, with red indicating the most recent interval. A color

legend was introduced at the top of the page for each visualization, e.g., see Fig. 2C for the disjoint intervals legend and Fig. 3F for the overlapping intervals legend.

2.2 The Proposed Visualizations

Figures 2 and 3 show the solutions based on disjunct and overlapping intervals, respectively. All examples use 4 typical categories of a news site, but categories can obviously be different and more than 4. The software that generates the visualizations from the number of PoI in each interval and category has been implemented by embedding calls to GD 2.0 (public graphics library that produces files in various formats, such as PNG and JPEG) into PHP scripts to allow for server-side dynamic image generation and inclusion into XHTML MP (Mobile Profile) pages.



Visualization A **Visualization B** **Visualization C**

Fig. 2. Visualizations based on disjunct intervals.

Visualization D **Visualization E** **Visualization F**

Fig. 3. Visualizations based on overlapping intervals.

Visualizations based on disjunct intervals (Fig. 2)**.** These visualizations highlight the relative proportions of the numbers associated to the 3 mutually exclusive intervals by using a single graphic element divided into 3 subparts, one for each interval.

The graphic element in *Visualization A* is an horizontal bar. The number in each subpart tells how many PoI have been added to the category during the corresponding interval. Width of bars always spans the whole screen, and width of subparts is proportional to the numbers, e.g. the 3 bars in Fig. 2A represent different situations where 50% of the updates have been made during the oldest of the 3 intervals.

Visualization B is similar to A, but employs pies instead of bars.

Visualization C employs stacked bar charts that refer to a common axis. Only the bar with the highest number of updates spans the whole screen, and it becomes possible to visually compare the width of bars among categories. The height of bars is smaller, so that more categories can be related on a single screen (also minimizing the replications of the reference axis to always have it displayed in case of scrolling).

Visualizations based on overlapping intervals. Reusing the previous visualizations also for overlapping intervals is not a good solution. Indeed, in the overlapping case, *Last 2h* contains *Last 20min*, and *Last 12h* contains both *Last 2h* and *Last 20min*, i.e. PoI associated to *Last 2h* include PoI of *Last 20min*, and so on. The previous visualizations show relative proportions of the numbers by dividing single graphic elements into 3 parts. Using them for overlapping intervals would produce charts where the *Last 12h* part would tend to fill most of the graphic element, making *Last 2h* and *Last 20min* visually disappear. We thus propose other visualizations (Fig. 3).

Visualization D employs a table: columns correspond to the 3 overlapping intervals and their colors, lines to categories; cells contain the number of PoI.

Visualization E employs a separate colored bar for each interval. Number of PoI is shown by text and by the width of bars. A bar spans the whole screen if it contains the highest number in its category. The 3 separate bars allow one to visually relate sizes inside a category and consider the inclusion relations that exist among intervals (e.g., from the business category in Fig. 3E, one notices that *Last 2h* and *Last 12h* coincide, i.e. PoI that arrived in the last 12h are precisely those that arrived in the last 2h).

Visualization F also employs 3 separate bars for the intervals, but draws bars with reference to a common axis (shown at the right of the page). It thus becomes possible to visually compare bars among categories. Since the usage of horizontal bars made it difficult to draw 4 categories in a single screen as we did in visualization C, we used here vertical bars so that more categories can be related on a single screen.

3 Conclusions

This paper motivated and proposed visualizations of thematic update status for sites aimed at mobile phone users. The next step in our research concerns a thorough evaluation of the proposed visualizations on users. In the remaining space, we can just briefly summarize the current main findings, i.e. (i) the results of the evaluation tend to encourage the use of visualizations based on overlapping rather than disjoint intervals, (ii) the presence of explicit numbers attached to each graphic element in some visualizations is another factor that proves to impact positively the results.

References

1. BBCi, http://www.bbc.co.uk/mobile/ (2004)
2. Financial Times, http://wap.ft.com (2004)
3. Kaikkonen A., Roto V.: Navigating in a Mobile XHTML Application. Proc. CHI 2003 Conf. Human Factors in Computing Systems. ACM Press, New York (2003) 329-336
4. MSN Mobile, http://mobile.msn.com (2004)
5. Nokia: Designing XHTML Mobile Profile Content, http://www.forum.nokia.com (2004)
6. Openwave Syst.: Best Practices in XHTML Design, http://developer.openwave.com (2004)
7. Passani, L.: Building Usable Wireless Applications for Mobile Phones. Proc. Mobile HCI 2002, LNCS 2411, Springer Verlag (2002) 9-20
8. The Wall Street Journal, http://wap.wsj.com (2004)
9. Tufte, E.R.: The Visual Display of Quantitative Information, Graphics Press (1982)
10. Yahoo Worldwide, http://wap.yahoo.com (2004)

Adapting Web-Pages for Mobile / PC Collaboration

Hui-Na Chua[1], Simon. D. Scott[1], See-Leng Ng[1], Michael Gardner[2], Anuroop Shahi[2], and Peter Blanchfield[3]

[1] BTexact Asian Research Centre, Cyberjaya, Selangor, Malaysia.
{hui.chua, simon.d2.scott, see.l.ng}@bt.com
[2] Chimera, University of Essex.
{mgardner, ashahi}@essex.ac.uk
[3] Division of Computer Science and IT, University of Nottingham Malaysia.
peter.blanchfield@nottingham.edu.my

Abstract. Web-based collaboration between mobile devices and PCs requires web-pages to be adapted to multiple devices. The paper begins by reviewing the considerations taken into account by existing web-content adaptation engines in adapting web-pages for the single-user browsing task. Next, the differences between single-user browsing and co-browsing are discussed along with the concept of shared view point and personal view point. Finally, a framework for adapting web-content for the purpose of co-browsing on different devices is outlined.

1 Introduction

With the introduction of web-enabled mobile devices, web-page designers face the challenge of ensuring that their web-content is displayed in a presentable fashion on a wide range of devices with vastly differing display capabilities. Two approaches have been developed to deal with this problem. The first approach is to create all of the many different versions of the same web-content, annotate them using the eXtensible mark up language (XML) and then specify which version of the content to display on a given device and its layout via the associated Stylesheet language (XSL). However this is a labor intensive task. A second approach is to dynamically alter the content retrieved by the web-server before it is displayed, usually through the use of proxy servers. A number of such automatic web-content adaptation engines have been developed by various groups [1, 2]. In these engines, the individual multimedia objects in a web-page are adapted through omission, summarization or conversion to a less resource intensive form. As the size of the display changes the layout of the multimedia objects also needs to change. Changing the layout requires semantic information to be extracted from the web-site; in other words the system needs to know how the different elements in the page are related and their functionality. Work in this area includes detecting the purpose of individual multimedia objects [3], determining how multimedia objects in pages are related to one another [4], and extraction of specific functional information from the entire web-site.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 421–425, 2004.

2 Single-User Browsing Content Adaptation Considerations

Adaptation is essentially a resource allocation problem where, subject to a set of constraints, the utility value of content presented in an adapted web-page is maximized. Whilst existing web-content adaptation engines largely apply the same set of constraints, namely display size, color-depth and ability to display certain types of web-objects such as Flash files; they differ in the considerations taken into account in the calculation of the utility of the content. These considerations can be split into the following categories:

- Relevance. Measuring the relevance of items of content has been done through click stream analysis [5] and determination of item purpose (such as advertisement, navigation, content or decorative) [3]. In most engines, irrelevant content is omitted completely and relevant content converted to less resource intensive forms.
- Informational content. As individual items are converted to less resource intensive forms it is often the case that the information content of the items is reduced and so this loss needs to be accounted for [6].
- Time. The quicker the content is displayed on the client device the better. Some adaptation engines estimate and factor in the time required to i) transform individual items of content to less resource intensive forms on the proxy server, and ii) uncompress content before display on the client device. To transform an individual item of content, the proxy server needs to download that item, which can introduce unacceptable delays especially if the bandwidth between the server and proxy or the proxy and client is much lower than that between the client and server [2]. Further, whilst compressed images are small and so result in bandwidth savings, the time required to compress an image on a heavily utilised proxy server and uncompress such image on a low end client device [7] can outweigh the benefits of doing so.
- Design Metrics. Design metrics are measures relating to composition (e.g. word count, link count), formatting (e.g. emphasized text, positioning) and other general characteristics (e.g. total bytes) of web-pages [8]. Scott and Koh demonstrated that for the single-user browsing task highly usable PDA web-pages have different design metrics compared to highly usable PC web-pages [9]. In other words, the presentation of web-content, such as amount of text emphasis and the number of colors used, needs to be considered in calculating the utility of the adapted content.
- Cost. Mobile telecom operators often charge GPRS / 3G users on the basis of the amount of data downloaded (kB), thus the cost of downloading adapted content items is an important factor in their comparative utility [10].

3 The Co-browsing Task

Web-based collaboration can be defined as two or more parties sharing sets of web-objects to pursue a common purpose. Normally this is achieved through co-browsing (also known as shared browsing and escorted browsing) where two or more users navigate a set of web-pages together from different clients whilst communicating with one another via an audio link or a text-chat application. Commercial co-browsing software, such as Microsoft's NetMeeting and Hipbone's Synetry CoBrowse, are

widely available and have been reviewed elsewhere [11]. However, such commercial software often makes the assumption that all the users are accessing the shared web-pages from a PC.

If a user is participating in the co-browsing session via a mobile device, then it is necessary to adapt the web-pages to that device. To ensure that the other participants can follow mobile user's description of the adapted web-page then it is necessary that they have a copy of the web-page as adapted for the mobile user on their own device. If there are multiple participants in the co-browsing session then all users should see a replica of the web-page as adapted for the smallest device, known as the Shared View Point (SVP). In addition to the SVP, those users on devices with larger displays will also be presented with their own Personal View Point (PVP) of the web-page in question. The PVP is the original web-page adapted to take into account the remaining available space on the device, the content already displayed within the SVP and the user's personal interests [12]. The reasoning behind this approach is that the users will refer to the SVP when discussing information, but will use the PVP to view other content from that web-page. The SVP can be changed by users i) dragging and dropping content from the PVP into the SVP, ii) following a navigation link in the SVP or PVP, iii) entering an URL into the address bar, or iv) downloading a bookmarked web-page. Fig. 1, shows a scenario where the user co-browsing via a PC has both a SVP and PVP, whilst the user of the mobile device sees only the SVP.

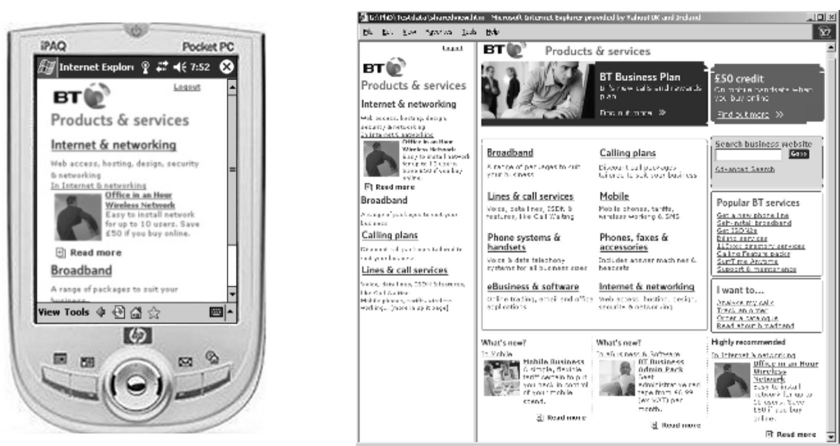


Fig. 1. Left - PDA with Shared View Point displayed in browser. Right - PC with both Shared View Point (left frame) and Personal View Point (right frame).

4 Adaptation Framework for Co-browsing

The proposed adaptation framework for co-browsing is shown in Fig. 2 below. The framework is divided into two parts (separated by a dashed line), namely the generation of the SVP followed by the generation of the PVP for each of the users. The scheme for deriving the actual layout of the SVP and PVPs from the relevant

content clusters is not shown since it would follow a similar scheme used by other adaptation engines in the literature [1, 9].

In generating the SVP, the majority of the considerations taken into account in calculating utility remain the same as in the single-user browsing task, but there are two changes. First, the capability of all the devices and the environments in which they operate must be considered in calculating factors relating to the time and cost of adapting items of content, as must the interests of all the users in the measuring the relevance of those items. Second, it has been demonstrated that the need to verbalize the content within web-pages has a significant impact on the usability of the web-pages and those web-pages with more visual cues, such as large graphics and emphasized text, are more usable [13]. In other words, the ideal design metrics of content adapted for the SVP are different from those adapted for the single-user browsing task. The only difference between generating the PVP and the adaptation of web-pages for the single-user browsing task is that the web-content present in the SVP needs to be account for. This is done through a reduction in the utility value of the content within the web-page that is already present in the SVP so as to decrease the likelihood that it would appear in the PVP unless there is adequate space.

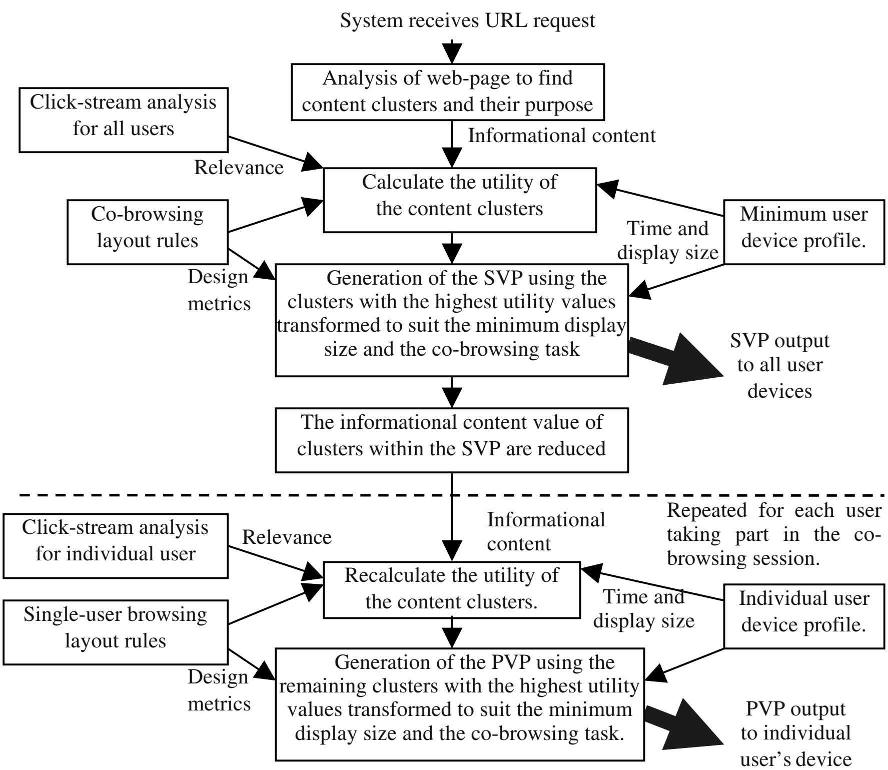


Fig. 2. Proposed Co-browsing Adaptation Framework

5 Conclusion

This paper has highlighted the considerations that need to be taken into account when adapting web-pages for co-browsing on different devices and a framework for doing so. Through informal experiments it has been found that i) the delay between the requesting new content and actually receiving it in its adapted form, and ii) the length of time between the first and last person receiving the SVP, have a large impact on the usability of content for co-browsing. Thus for practical usage it is envisaged that all users on mobile devices will see only the SVP, whilst users on PCs will see both the SVP and original page in its unadapted form.

References

[1] Bickmore, T., Girgensohn, A. and Sullivan, J.W. Web Page Filtering and Re-Authoring for Mobile Users. In the Computer Journal, 42(6), pp 534-546, 1999.

[2] Ma, W.Y., Bedner, I., Chang, J., Kuchinsky, A. and Zhang, H.J. A Framework for Adaptive Content Delivery in Heterogenous Network Environments. Proc of SPIE / ACM Conference on Multimedia Computing and Networking, Feb 2000.

[3] Paek, S., Smith, J.R. Detecting image purpose in World-Wide Web documents. In Proc. IS&T / SPIE Symposium on Electronic Imaging: Science and Technology – Document Recognition, San Jose, CA, 1998.

[4] Chen, J., Zhou, B., Shi, J., Zhang, H.J. and Fengwu, Q. Function-Based Object Model Towards Website Adaptation. Proc. of 10th Int. WWW Conference, 2001.

[5] Anderson, C.R., Domingos, P. and Weld, D.S. Personalizing Web Sites for Mobile Users. Proc. of 10th Int. WWW Conference, 2001.

[6] Mohan, R., Smith, J.R. and Li, C.S. Adapting Multimedia Internet Content for Universal Access. IEEE Transactions on Multimedia, pp 104-114, March 1999.

[7] Han R. Factoring a Mobile Client's Effective Processing Speed Into the Image Transcoding Decision. In 2nd ACM International Workshop on Wireless Mobile Multimedia (WOWMOM), pp 91-98, 1999.

[8] Ivory, M.Y., Sinha, R.R. and Hearst M.A. Preliminary Findings on Using Quantitative Measures to Compare Favorably Ranked and Unranked Information-centric Web Pages. In Proc. of the 6th Conference on Human Factors and the Web, 2000.

[9] Scott S.D. and Koh Y.H. Design Metrics and the Adaptation of Web-page Content Chunks for PDAs. In Proceedings of the 3rd Internation Conference on Information Technology in Asia (CITA03), pp 261-268, 2003.

[10] Kylänpää, M. and Heino, I. Personalized Transcoding Proxy - an Approach to Mobile Web Access. In Proc. of 9th Int. WWW Conference, 2000.

[11] Esenther, A. W. Instant Co-Browsing: Lightweight Real-Time Collaborative Web Browsing. In Proc. of 11th Int. WWW Conference, 2002.

[12] Chen, C.C., Chen M.C. and Sun Y.S. PVA: A Self-Adaptive Personal View Agent. In Journal of Intelligent Information Systems, Vol. 18, pp 173-194, 2002.

[13] Chua, H.N., Scott, S.D., Tham, V.K., Gardner, M. and Blanchfield, P. Design Metrics for Co-browsing. In 7th International Conference on Working with Computing Systems (WWCS2004), 29 June – 2 July, Kuala Lumpur, 2004.

EMG as a Subtle Input Interface for Mobile Computing

Enrico Costanza, Alberto Perdomo, Samuel A. Inverso, and Rebecca Allen

Liminal Devices, Media Lab Europe
Sugar House Lane, Bellevue, Dublin 8, Ireland
{Enrico.Costanza, Alberto.Perdomo, Sam.Inverso, Rebecca.Allen}@medialabeurope.org
http://www.medialabeurope.org/

Abstract. Rather than merely imitating the desktop metaphor for mobile devices, new interface paradigms that take into account the particular characteristics of mobility, need to be developed. In this paper an input device based on the electromyographic (EMG) signal is proposed as a controller for mobile interaction. The interface can be considered subtle or intimate because individuals are able to interact privately without causing distraction to their immediate environment. The results from a preliminary study are presented to show the feasibility of the proposed system.

1 Introduction

With recent advances in microelectronics and display technology, current handheld devices such as mobile phones and PDAs are now powerful computing platforms that support network connectivity and embed colour touch screens. The user interfaces for these devices are generally derived from graphical interfaces for desktop computers using reduced versions of the keyboard and mouse. Rather than merely imitating the desktop metaphor for mobile devices new interface paradigms that take into account the particular characteristics of mobility, need to be developed. In a mobile context the user's attention should not be totally or even largely devoted to the computer interface. In addition, consideration should be given to the form of interaction in relation to the type of tasks that can be carried out in a mobile environment and its social acceptance. When the user is on the move or engaged socially, most of the computer interaction is of short duration. Often the user will be involved in simultaneous activities, e.g. talking, walking, and may be just checking for incoming messages in his mailbox.

A partial solution to the problems mentioned can be found in the use of output forms like audio [1,2] haptics [3] or graphic displays embedded in eyeglasses [4,5]. Different forms of input and output should be integrated in a multimodal interface to adapt to different tasks and situations. However, the interaction design for this type of systems constitutes an open challenge: the ideal mobile device should be 'hands-free' and 'eyes-free'.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 426–430, 2004.

In this paper an input device based on the electromyographic (EMG) signal is proposed as a controller for mobile interaction. EMG allows the sensing of intentional muscle activity not necessarily related to articulation. In this way a class of *"motionless" gestures* can be defined to control applications on mobile devices. Such an interface can be considered subtle or intimate because individuals are able to interact privately without causing distraction to their immediate environment. This may improve the social acceptance of the interface.

2 Related Work

The electromyogram is an electrical signal generated by neuromuscular activity [6]. It can be recorded non-invasively using surface electrodes. Methods for effective recording and computer aided analysis of EMG signals have been the object of study in the field of biomedical engineering for the last three decades. EMG signals have been modelled as Gaussian like coloured zero mean noise [7]. Typical biomedical analysis involves envelope detection, energy measurement (directly related to the force) and frequency characterization. Research in this domain focuses on diagnosis.

Other studies in the domain of bioengineering have concentrated on the use of electromyographic signals for control of prosthesis, rehabilitation and computer interfaces for users with motor disabilities [8,9]. Beyond medical applications, EMG has been proposed for control of computer interfaces. Examples include interfaces for musical expression [10], controls for consumer electronics [11] and videogames [12]. All of these are based on EMG signals acquired from the forearm.

EMG based interfaces generally involve signal acquisition from a number of differential electrodes, signal processing (feature extraction) and real-time pattern classification. Classification methods based on both statistical and neural network approaches have been reported with satisfactory results. However, given the complexity of the task and the variability of the EMG signals [13] these systems usually require calibration for each user or training of the pattern recognition algorithms.

In a different fashion, but still in the context of HCI, EMG signals have been used in conjunction with other physiological signals (skin conductivity, blood pressure and respiration) to detect the affective state of the user [14].

A number of input interfaces based on gestures have been proposed for mobile and wearable computing. The most common approach is based on inertial sensors (accelerometers) worn by the user [1] or included on a PDA/mobile phone [15,16]; Rekimoto presented an interesting alternative based on capacitive sensing [17]. Some of the studies pose questions related to the social acceptance of the proposed gestures.

3 Concept

EMG can be used to sense isometric muscular activity [18]: the type of muscular activity that does not translate into movement. This feature makes it possible to define a class of *subtle motionless gestures* to control an interface without being noticed and

without disrupting the surrounding environment. A simple generic controller in the form of an ancillary device is proposed. It can be placed on a muscle (for example the bicep) and activated by its contraction. When activation is detected, the controller sends a signal wirelessly to the main wearable processing unit, such as a mobile phone or PDA. The device is attached to an adjustable elastic band that can be hidden under clothes. Surface electrodes are integrated on the inside to acquire the signals. These are amplified, filtered and processed by integrated components. Compared to other EMG based controllers the approach proposed is to trade resolution (in terms of number of different gestures being recognized) for robustness and eliminate the need for calibration, yet keeping the computational cost to a minimum so that no external processing is required.

This simple controller can be used within a multimodal interface. In an example scenario the system has a display (eyewear or audio) capable of delivering high resolution information such as text (requiring a certain level of attention), as well as delivering low resolution *peripheral cues* (that do not require as much attention). Events such as new messages or phone calls generate cues. The user can react to cues by contracting the muscle, for example requesting more information about the event (e.g. the message subject or the caller ID). The peripheral cues can otherwise be ignored, if the user cannot afford to give attention to the computer. Using EMG, the user can react to the cues in a subtle way, without disrupting their environment and without using their hands on the interface.

4 Preliminary Study

A preliminary study was carried out to explore the feasibility of the wearable EMG controller. A prototype was developed to record EMG data from moving subjects. The device acquires the physiological data and streams it wirelessly to a PC used for logging and offline analysis of the signals.

Three surface electrodes (input, reference and ground) are used to acquire the signal from the muscle. The input and reference signals are connected to an instrumentation amplifier and then filtered using a high-order low pass filter. An 8-bit microcontroller equipped with an integrated analogue to digital converter is used to sample the signal and transmit it to the PC using a Bluetooth™ module.

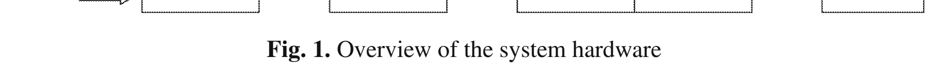


Fig. 1. Overview of the system hardware

A group of 10 subjects (6 males, 4 females) between 25 and 33 years of age took part in the study. The subjects were informed of the purpose of the study and the function of the controller and then asked to "contract their muscle" in reaction to an audio stimulus. A total of 10 stimuli were presented aperiodically to each subject

within 4 minutes. The sound was synchronized with the data logging, to facilitate analysis. The electrodes (Ag/AgCl) were placed on the bicep and the device was worn on an armband. During the test the subjects were asked to move freely within a range of 10 meters to simulate realistic conditions.

A simple algorithm was designed to detect brief muscle contractions in the recorded EMG signal. The starting point for the design was the observation that in correspondence to a short muscle contraction the signal exhibits a peak, and that the duration of the peaks appeared to be consistent (≈ 0.60 - 0.80 s) across the different subjects even if no precise instruction on the duration of the contraction had been given. The signal was rectified and filtered with a moving average low pass filter tuned on the peak duration. Peaks are detected according to the following two conditions:

$$x_n - x_{n-T} < K_1$$

$$K_2 < x_{n-T} - \frac{1}{4}\left(\sum_{i=1}^{4} x_{n-(i+1)T}\right)$$

where x_n denotes the current sample and T corresponds to a delay of 0.75s. The values of K_1 and K_2 were obtained by training the algorithm on a subset of the total data acquired. The training consisted in minimizing false positives and false negatives on the data collected from 4 of the subjects.

The muscle contraction was correctly detected in 84% of all cases (10 subjects). A number of 126 false detections occurred over the 40 minutes of signal recorded (running the algorithm on all the available samples), independently of the number of stimuli. In a realistic scenario the software would look for a peak only after a cue is presented to the user. Hence the probability of a *false trigger* corresponds to the probability of a false peak detection occurring right after a cue to which the user does not want to respond. In this case, the number of *false triggers* would be considerably smaller than the number of false positives.

5 Conclusion

A *subtle* EMG based controller for mobile computing has been proposed. Results from a preliminary study show that even with simple processing techniques it is possible to detect brief muscle contractions in data acquired from moving subjects.

The results encourage further development of the interface. The signal processing and pattern recognition strategies should be improved to achieve higher accuracy. At the same time, the efficiency of the interface can be increased introducing feedback. The use of dry electrodes is being considered to promote user acceptance. Other muscles beside the bicep will be considered, including the combination of different ones. More in general, the authors plan to study the integration of the controller within a multimodal interface and the interaction design for the mobile context. Applications should be developed and user studies conducted to validate the general usability.

Acknowledgments. The authors would like to thank the MindGames group, Matt Karau and Renee Hall at Media Lab Europe for their precious support.

References

[1] Lumdsen, J., Brewster, S.: A Paradigm Shift: Alternative Interaction Techniques for Use with Mobile & Wearable Devices. In: Proceedings of 13th Annual IBM Centers for Advanced Studies Conference CASCON'2003, Toronto, Canada.

[2] Sawhney, N. Schmandt, C.: Nomadic Radio: Scaleable and Contextual Notification for Wearable Audio Messaging. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, Pittsburgh.

[3] Toney, A., Dunne, L., Thomas, B., Ashdown, S.: A Shoulder Pad Insert Vibrotactile Display. In: Proceedings of the 7th IEEE International Symposium on Wearable Computers (ISWC 2003), White Plains, New York.

[4] http://www.microopticalcorp.com/

[5] http://www.eyetap.org/

[6] De Luca, C.J.: Physiology and Mathematics of Myoelectric Signals. In: IEEE Transactions on Biomedical Engineering, Vol. BME-26, No.6, June 1979.

[7] De Luca, C.J.: Surface Electromyography: Detection and Recording. Whitepaper, DelSys Inc., 2002.

[8] Yamaguchi, H., Nishikawa, D., Yu, W., Maruishi, M., Yokoi, H., Mano, Y., Kakazu, Y.: EMG Control Switching System for FES. In: Proceedings of the 1999 IEEE International Workshop on Robot and Human Interaction, Pisa, Italy.

[9] Moon, I., Lee, M., Ryu, J., Mun, M.: Intelligent Robotic Wheelchair with EMG-, Gesture- and Voice-based Interfaces. In: Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, NV.

[10] Tanaka, A., Knapp, R.B.: Multimodal Interaction in Music Using the Electromyogram and Relative Position Sensing. In: Proceedings of the 2002 Conference on New Instruments for Musical Expression (NIME-02), Dublin, Ireland.

[11] http://www.csl.sony.fr/~atau/gesture/

[12] Wheeler, K.R., Jorgensen, C.C.: Gestures as Input: Neuroelectric Joysticks and Keyboards. In: IEEE Pervasive Computing, April-June 2003 (Vol. 2, No. 2).

[13] Dubost, G., Tanaka, A.: A Wireless, Networked-based Biosensor Interface for Music. In: Proceedings of the ICMC 2002, Göteborg, Sweden.

[14] Healy, J., Picard, R.: Digital Processing of Affective Signals. In: Proceedings of ICASSP, Seattle, WA, 1998.

[15] Ängeslevä, J., O'Modhrain, S., Oakley, I., Hughes, S.: Body Mnemonics. In: Proceedings of Physical Interaction (PI03) – Workshop on Real World User Interfaces, a workshop at the Mobile HCI Conference 2003, Udine, Italy.

[16] Lehikoinen, J.: Virtual Pockets. In: Proceedings of the Fourth International Symposium on Wearable Computers (ISWC'00), Atlanta, Georgia, 2000.

[17] Rekimoto, J.: GestureWrist and GesturePad: Unobtrusive Wearable Interaction Devices. In: Proceedings of the 5th IEEE International Symposium on Wearable Computers (ISWC 2001), Zurich, Switzerland.

[18] Greenman, P.E.: Principles of Manual Medicine. Publisher: Lippincott Williams & Wilkins Publishers.

Mobile Support for Team-Based Field Surveys

Malcolm Hall and Philip Gray

Department of Computing Science, University of Glasgow, Glasgow, G12 8QQ, United Kingdom
{mh, pdg}@dcs.gla.ac.uk
http://www.dcs.gla.ac.uk

Abstract. This paper describes a study of the use of multimedia networked location-aware mobile computers to support team-based survey-oriented fieldwork. Existing systems do not provide fully integrated support for collaborative data capture and review, or access to distributed real time information on survey progress and status, all of which are crucial for the conduct and management of surveys often carried out under inflexible time constraints. We developed a mobile system to address these shortcomings and performed an evaluation in an archaeological field survey, supporting over two-hundred data collection incidents over five days, and providing further insight into the field work data collection process.

1 Introduction

Mobile computers, wireless networking and positioning technologies are becoming increasingly suitable for outdoor work. However, determining how to use them effectively in team environments still remains a challenge. Work by Fagrell *et al* [1], and a number of commercial systems, demonstrates the value of mobile computers supporting teams of outdoor workers; particularly for collaboration and the coordination of activities. Team-based field surveys, such as archaeological field studies, are a promising application area for these technologies, since they involve intense collaboration over a distributed area. In team-based field surveys two essential requirements are positional awareness and team coordination. Mobile technologies can not only meet their current requirements, but potentially offer unique, previously unrealised benefits through the real time update of information on fieldwork progress supporting timely coordination of the team effort.

This paper describes our proposal of a multimedia oriented, location aware system to support team-based field studies. This system allows field workers to share their position and activity with other fieldworkers, as well as collecting photographs and data at their current location, and share this wirelessly with other fieldworkers. We present the results of an initial requirements capture involving professional archaeologists, and the resulting prototype system addressing their needs. Finally, we describe an evaluative field trial of the system and conclude with observations on the potential of this technology.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 431–435, 2004.

2 Supporting Team-Based Field Work

Pascoe *et al* [2] defined four characteristics of the mobile field worker – "*dynamic user configuration, limited attention capacity, high speed interaction and context dependency*" and subsequently developed the Minimal Attention User (MAU) interface. Fieldnote [3] is their proposed system featuring a MAU interface which utilises a database to allow the collation of users' data. While innovative in terms of its interface, the Fieldnote system has no real-time synchronous support for teamwork. Other systems such as Renevier and Nigay's [4] 'MAGIC' system, and RAMSES [5] have similar shortcomings in that they focus on single user data collection and fail to support synchronous collaborative fieldwork.

The importance of team work in archaeology became apparent to us in our early design studies of archaeological work. We started by examining a typical field survey, part of the South East Melos Project (performed by our University's Archaeology Department) whose aim was to locate evidence of early mines and Roman quarries. Through interviews and observations we were able to outline the areas where computer support would be beneficial and identified some key requirements.

In the typical archaeological field survey we studied, several small teams are each assigned to a particular sector of the survey area and responsible for recording and investigating all features of interest. Each team needs to be aware of its location in its assigned sector and should be able to create records, visual or otherwise, of its findings. Post-survey the data needs to be gathered and organised into an easily accessible format to facilitate further study.

To assist location awareness the Melos team used handheld GPS devices, a compass, an aerial photo and a map. It was noted that using the equipment together was a time-consuming and laborious process, further aggravated by the apparent incompatibility between the area map coordinate system provided and those supported by their GPS units. This suggests the possibility of support for automatically overlaying GPS information onto a geo-referenced digital map.

The creation and cataloguing of records was addressed by the Melos team using paper forms and a digital camera. This proved problematic as the task of physical record management was left to the user. In fact, it was often the case that paper forms were misplaced or lost by a surveyor during an expedition. Finally, management and data entry of gathered records was originally performed manually by project leaders in the Melos group through a lengthy and error-prone process.

3 System Design

To support these collaborative activities during field work our system uses a client-server architecture. The server is designed to be deployed on-site in order to coordinate data exchange among the client devices, and act as a central record repository. The client is designed to operate on several lightweight mobile devices, with limited processing abilities. In practice we expect our system to be used with a single server and 5-10 client devices, one for each team, though provisions have been made for scalability.

The client is a Java application which provides a map viewer and record entry components. The map viewer (see left side figure 1) displays a geo-referenced map image and supports panning and zooming using a stylus or the directional button. The user's current location is displayed as a marker, and is broadcast to the other users, allowing all users' locations to be displayed. Similarly, markers are used to display the location at which records were created, and this information is distributed to all users. The map displays polygons marking the area each team has been allocated to survey, and also indicates the area covered. These features aid orientation and provide easily accessible insight on a team's progress. Other features enhancing usability include a distance measurement tool and the ability to re-position previously recorded observations on the map.

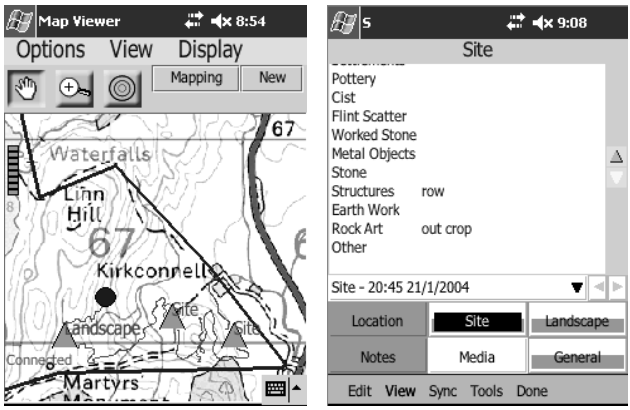


Fig. 1. The map and form interfaces to the client application

Record entry is handled via the form view (see right side figure 1). Each record entry is divided into sections which are accessible using the panel at the bottom. The buttons are colour-coded to convey each section's completion status.

The sections are as follows: a GPS section which is automatically filled when the record is created; a media section which displays the camera viewfinder and allows the attachment of photos and voice recordings; structured input sections for the records contents; and a section for general notes. The form structure is flexible as it is specified in XML and can be easily altered using a form builder application. We used KelvinConnect's KC form engine which is based on the Paraglide system [6] and utilises optimized input techniques, such as pick lists and auto-completion.

The server component of the system is implemented as a Tomcat Servlet using an XML database, and includes an integrated FTP server for managing multimedia content. The server provides support during and post survey. During the survey the server receives and distributes records and GPS locations from each client. If the user drifts out of network coverage the recorded data is cached and flushed to the server when the link is resumed.

For post-survey services the XML records can be converted into a variety of formats including HTML, allowing the automated creation of a web site which can display categorised coverage maps of various findings, for example to display a colour coded map of the areas where features of a certain age were found.

4 Evaluation

The final evaluation of the system took place during a five day archaeological survey in Kirkcudbrightshire, Scotland. The purpose of the survey was to record the GPS location of all known and suspected sites containing features of interest, from the Mesolithic to the Middle Bronze Age. As the survey area was much larger than the wireless range of the equipment, the server was deployed on a laptop acting as a portable access point.

Fig. 2. The system on an iPAQ equipped with camera and GPS

We used four HP iPAQs with integrated 802.11b, GPS and digital camera. Figure 2 displays the client configuration as used in the trials. This compact, lightweight configuration possesses most of the functionality that field workers usually require and can be held comfortably in the hand.

The expedition was successfully completed within the allotted time without any serious problems. For half the expedition time, the participants were asked to use their normal techniques and practices for the survey and for the remaining time to utilise the new facilities provided by our system. During the five days of the trial, over two-hundred record entries were made. While the system used was only a prototype it successfully supported collaborative field work. In particular, fieldworkers could easily see which sections of the area had already been surveyed, avoiding the duplicated work we had observed as part of the Melos project.

One problem with the system concerned familiarisation, since many of the surveyors were not experienced with using handheld technology. Another issue discovered was the intangible nature of the data collecting process; users were not completely satisfied with not having hard copies of their reports at hand, even though their observations were cached on their palmtops and readily available at any time.

However, the most experienced surveyors in the group provided positive comments on the system as it automated many tasks which previously were performed manually. The most useful of these features, identified by the users, was the GPS positioning on the map which reduced the effort required for navigation and automated record location entry.

While the short nature of this trial ruled out collecting extensive data on the effectiveness of the system, we believe that a noticeable increase in productivity

would be apparent after prolonged use and familiarisation with the system. In our future work we plan to test this system over a longer period of time to judge its effectiveness with respect to existing paper based technologies.

5 Conclusion

We developed a system to facilitate collaboration and coordination in large field surveys. Our approach is unique in focus, realising effective team work whilst replicating the useful functionality of common surveyor tools. To evaluate the system we put it to use in an actual archaeological survey and received positive feedback.

In the future we aim to expand on the systems decentralised nature by implementing part of the server's functionality on the client side. Although a central repository is still useful, by decentralising the system it is possible to exploit the potential of ad hoc networks to greatly expand the coverage area of the system. Finally, it would be interesting to note and observe in practice the potentially wide applicability of such a tool in other scientific domains such as geographical or environmental surveys.

References

1. H. Fagrell, K. Forsberg, and J. Sanneblad. *FieldWise: a mobile knowledge management architecture*. in Proceedings of the 2000 ACM conference on Computer supported cooperative work. 2000: ACM Press.
2. J. Pascoe, N. Ryan, and D. Morse, *Using while moving: HCI issues in fieldwork environments.* ACM Transactions on Computer Human Interaction, 2000. **7**(3): p. 417-437.
3. N. Ryan, J. Pascoe, and D. Morse, *FieldNote: extending a GIS into the field*, in New Techniques for Old Times: Computer Applications in Archaeology, 1998, J.A.Barcelo, I.Briz, and A.Vila, Editors. 1999.
4. P. Renevier and L. Nigay. *Mobile Collaborative Augmented Reality: the Augmented Stroll*. in EHCI. 2001. Toronto.
5. M. Ancona, G. Dodero, and V. Gianuzzi. *RAMSES: a mobile computing system for field archaeology*. in Handheld and Ubiquitous Computing. 1999. Karlsruhe.
6. M. Gardner, M. Sage, P. Gray, and C. Johnson. *Data Capture for Clinical Anaesthesia on a Pen-based PDA: Is It a Viable Alternative to Paper?* in HCI. 2001. Lille.

"Please Turn ON Your Mobile Phone" – First Impressions of Text-Messaging in Lectures

Matt Jones[1] and Gary Marsden[2]

[1] Dept of Computer Science, University of Waikato, Private Bag 3105, Hamilton, New Zealand, always@acm.org
[2] Dept of Computer Science, University of Cape Town, Private Bag Rondebosch 7701, Cape Town, South Africa, gaz@cs.uct.ac.za

Abstract. Previous work by Draper and Brown [3] investigated the use of specialized handsets to increase interactivity in lecture settings. Inspired by their encouraging findings we have been exploring the use of conventional mobile phones and text-messaging to allow students to communicate with the lecturer as the class proceeds. In our pilot-study, students were able to respond to MCQs and send free-text comments and questions to the lecturer via SMS. Through observations and interviews with students and lecturers, we gained useful impressions of the value of such an approach. Students enjoyed the opportunity to be more actively involved but voiced concerns about costs.

1 Introduction

Anyone who has given a talk or lecture to a large audience will be well-acquainted with the uncomfortable silences, embarrassed glances and nervous shuffling that greet requests for audience participation. This anecdotal evidence is supported by survey findings presented by Draper & Brown [3] indicating that if a lecture class is asked for a verbal response, 0 to 3.7% of students is likely to respond; even for the less exposing, "hands-up" response style, the participation rate is also a low 0.5-7.8%.

Not all audiences are so shy, though. In the late-1990s the television game show, *"Who wants to be a millionaire?",* attracted large viewing numbers throughout the world. As part of the game format, the contestant could "ask the audience", getting each member to answer the multi-choice question using a handset.

Draper and Brown have taken similar handsets out of the TV studio and into the classroom. In [3] and an earlier paper [2], they present pedagogic motivations for their work which we share and will not elaborate on here beyond noting the value of interactivity and engagement between the learners (students) and the learning-leader (lecturer).

In a long-term, extensive study – summarized in [3] – the personal response system they used for multiple-choice questions (MCQs) was seen as being of benefit: for example, 60% of 138 first-year computer students rated the system "extremely" or "very" useful; and, similar responses were seen in other disciplines as varied as medicine and philosophy. Handsets are also likely to increase the participation levels – when asked whether they would work out an answer if asked to vote using the system, between 32-40% agreed.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 436–440, 2004.

Of course, specialized handsets have many advantages such as providing simple, direct ways for students to respond (they just press a button); however, there are some drawbacks including: large costs involved in providing handsets ubiquitously, for every student and every lecture; organizational-overheads (e.g. handing out and collecting handsets); and, the impoverished range of responses possible (a single selection for MCQ use).

Inspired by Draper and Brown's experiences we sought to address these sorts of drawbacks by using a technology that most developed-world students now carry with them to every lecture – the mobile telephone. We were interested in whether the pervasiveness and easy familiarity students have with this technology would allow it to serve as a replacement for the purpose-built handsets. Furthermore, we wanted to explore the possibilities beyond MCQs such as students sending free text questions or, perhaps suggestions and comments to the lecturer. Although other researchers have considered the use of mobile phones in a university setting (e.g., [1]), we believe this to be a novel application.

2 Example Scenario

While the specialized handset studies provided us with a very useful set of functional and non-functional possibilities, we decided to also run some sessions bringing together a group of eight experts in both human-computer interaction and education (all of which were also lecturers) to brainstorm requirements. In the process we developed scenarios such as this one:

Dr Monday begins her lecture on advanced linguistic analysis to 300 first year students. "Before we go any further, are there any questions about last week's topic? Send me a text now from your mobile phone to 444". After a minute, Dr Monday checks the computer display and sees there are 25 questions listed in the order they arrived; she can reorder the list alphabetically and by size of message as well. She selects one of the questions to answer.

Later in the lecture, Dr Monday wants to test the students' understanding of "focus". "Here's a quick quiz," she says. "If you think focus is related to the subject, text 1 to 444; if you think it is related to the topic, text 2; and if you think it is related to the verb, text 3 to 444". Moments later, Dr Monday can display a bar chart showing the students what the most popular choice was. "Most of you are wrong", she says, wryly, "the correct answer is 2 – the topic".

Several times in the lecture, Monday asks the students to text their current "happiness level": "send a text message to 444 now to show how well you understand the lecture so far," she says, "enter H followed by a number from 0 to 9 where 0 is the worst". She can view the changing level of "happiness" over time as a line graph.

After the lecture, Monday returns to her office and can access all the questions sent by students; she can also review the bar charts for each multiple choice question; and see the "worm" trace plotted over time. All this information helps her review the lecture content and plan for next week's session.

Such discussions clarified some of the additional forms of interactivity mobiles might provide over specialised handsets: allowing multiple responses to a MCQ – e.g., "choose 2 of the 5 features listed below"; parameterised responses – e.g. "text your answer (1-5) and how confident you are in your answer (0-100%)"; open-ended 'conversations' between the lecturer and audience; and, finally, as an active feedback device.

3 Pilot-Study System

Before building a full-scale system, tailored specifically to the lecture-context, we decided to acquire a third-party, commercial text-polling system to first explore the issues and feasibility of our ideas. The software chosen was the *SMS PollCenter* by Code Segment[1]. The system runs on a PC (we ran it on a laptop in the field studies) and also requires a mobile phone to be connected to the computer via a serial cable so that sent text messages can be gathered. MCQ results can be displayed in a range of forms such as bar chart and a pie-chart. The "SMS Chat" facility displays incoming texts in a scrolling whiteboard format.

4 Field Studies

We studied the system in use over six, one-hour sessions spread over a couple of months. Our aim was to gather impressions in a range of contexts so we chose situations with different characteristics and used the system in a variety of ways.

Three courses were involved: *A*- first year programming class run in New Zealand (NZ); *B*- first year programming class run in South Africa (SA); and, *C*- a 4th year human computer interaction class in South Africa. For courses *B* and *C* we carried several trials each separated by around a week. During each session, researchers set up and operated the system for the lecturer; they also observed the class interaction and were involved in interviewing students at its end. In class *A* and *C* the authors were the lecturers – we wanted to experience the system from the front, as it were; two other lecturers were involved in presenting class *B*.

A summary of each session and use of the system within them is shown in Table 1, along with data on the number of text messages received during each use. While this table gives some raw indications of interactivity, it is worth highlighting some of the specific behaviours and effects we noticed. First, 19% of all logged responses to MCQ style questions were in a form that were not recognized by our answer matching filters: for example, in Session 2.1, the students were asked to enter a single integer, but one sent *"Turn 72 degqees"* (*sic*). Second, on average, 10% of respondents sent more than one message in response to a question (either resending their initial response or changing their vote). Third, in SA, 6% of all messages were spam (e.g., *"Let the universe decide SMS "oracle" to 34009"*); no spam was received in NZ. Fourth, in most of the MCQ cases, as the lecturer discussed the results of the poll

[1] For information and a demonstration see: http://www.codesegment.com/

Table 1. Summary of sessions and system use. In each session (e.g. 2) there was one or more use of the system (e.g. 2.1, 2.2). Questions were either factual (based on lecture content) or personal (eliciting subjective opinion). Text messages sent were either single selections relating to a MCQ or free text (chat style). Messages/poll results were either were fully visible (results shown during polling and dynamically updated), partially visible (final results shown at end of polling) or hidden (only the lecturer saw the messages).

Session/ system use	Course	Question type	Response elicited	visibility	# people in class	#unique respondents (% of total)
1	A	factual	MCQ	full	155	35 (23%)
2.1	B	factual	MCQ	full	180	32 (18%)
2.2	B	personal	chat	full	180	16 (9%)
3.1	B	personal	MCQ	partial	150	17 (11%)
3.2	B	factual	MCQ	partial	150	10 (7%)
4.1	C	personal	MCQ	full	40	15 (38%)
4.2	C	personal	chat	full	40	3 (1%)
5.1	C	factual	MCQ	full	40	6 (15%)
5.2	C	personal	chat	hidden	40	3 (1%)
6.1	C	personal	MCQ	full	33	10 (30%)

chart, additional messages would arrive – sometimes this was a mobile telephone network effect (5-10% of messages were delayed), but there was also evidence of a 'playfulness' as students attempted to 'disrupt' the lecturer by altering the results.

At the end of each session, we asked for volunteers to remain behind and give feedback on the system. Overall we spoke to around 50 people in this way. Views were consistent in that students liked the idea of the approach (it gave them more of a role in the lecture, changed the pace of the session *etc*); strongly preferred the MCQ style of interaction over the chat scheme (as texting a freeform question could take too long and the display of comments to the whole class could be distracting); but, they had concerns over the cost of sending messages (over and over again we were told – *"if sending a message was at a reduced rate, or free, I'd use it a lot more").*

We also discussed the experience with the class *B* lecturers. They were less enthusiastic and more cautious about the scheme than the students. Their main concerns were the potential negative impacts of the technology on the "natural" flow of the lecture and the need for more flexibility in the software to respond dynamically.

5 Discussions and Future Work

As this was a pilot-study, no strong conclusions can be drawn at this stage. However the results suggest that using the handsets to SMS responses to MCQs could improve the level of participation: we saw a response rate of 7%-38% (much higher than that predicted by Draper and Brown for 'hands-up'). The system was most successful when the results were always on display to the students (from the start to the end of the poll): we discovered that students liked watching their messaging change the display dynamically. Even when the messaging rate was low, the technique appeared

to have a positive impact on the lecture experience: the sessions became more participative with the lecturer engaging the students in a discussion of the poll results, for instance. In setting up software to process MCQ responses, the aim should be to accommodate the variety of answer messages likely to be sent (e.g. “1”, “one”, “the first choice”).

While a novelty effect might well have been in play, the response rate seen in 6.1 (30%) compares favorably with that in 4.1 (38%), even though the second session took place approximately one month after the earlier one. Given Draper and Brown’s experience, we predict the enthusiasm for the approach would grow, particularly if charging issues can be resolved (e.g., by providing free texting for students).

The ‘chat’ form of interaction was disappointingly received. However, we intend to explore this form further as its potential was undermined by the constraints of the pilot system (e.g. lack of filtering or censoring facilities for the lecturer). Another area for potential was discovered in the form of interesting emergent ‘community’ behaviour when the chat screen was visible to all students: as well as communicating with the lecturer, students posed questions to *each other* and received replies from within the audience. While there is much exciting work on mobile communities for non-collocated people, this experience suggest there is some useful work to be done on supporting *immobile* mobile communities, such as crowds in football stadia.

Acknowledgements. Thanks to Hussein Suleman and Donald Cook who set aside time in their lectures. Dave Nichols and Mike Mayo helped with the NZ observations and the Waikato HCI group worked on scenarios. Steve Draper gave useful comments on a earlier draft.

References

1. Cheverst, K. Dix, A., Fitton, D., Friday, A. and Rouncefield, M. Exploring the Utility of Remote Messaging and Situated Office Door Displays, in *Proceedings of MobileHCI '03*, LNCS 2795, Springer.
2. Draper, S. W., Cargill, J & Cutts, Q. (2002). Electronically enhanced classroom interaction”. *Australian Journal of Educational Technology* vol. 18. no. 1, pp13-23.
3. Draper, S. W. & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning* 20, pp81-94. Blackwell.

A Stereoscopic Image Rendering Method for Autostereoscopic Mobile Devices

Daehee Kim, Hoonjong Kang, and Chunghyun Ahn

Electronics and Telecommunications Research Institute (ETRI)
161 Gajeong-dong Yuseong-gu, Daejeon, 305-350, Korea
{daeheekim, khj63409, hyun}@etri.re.kr

Abstract. A novel stereoscopic image rendering method for autostereoscopic mobile devices is presented in this paper. The system to implement the proposed method consists of autostereoscopic display system and the stereoscopic rendering software engine for mobile devices. First, we mount a parallax barrier, which is made by a twisted nematic (TN) panel, on the liquid crystal display (LCD) to display stereoscopic 3-D images generated by the proposed stereoscopic rendering software engine. In addition, we present the stereoscopic rendering algorithm for 3-D graphic models. The proposed algorithm generates left-view images and right-view images from 3-D graphic models. Therefore, the proposed rendering system provides autostereoscopic views in order that users can enjoy three dimensional effects without any special glasses.

1 Introduction

Recently, various contents for mobile devices have been focused to accommodate users' various demands in mobile markets. In addition, mobile markets are interested in representing 3-D computer graphic data on mobile devices. However, since conventional 3-D rendering methods are based on the monocular view, it is difficult to provide sensation of reality to users. In order to alleviate this problem, stereoscopy can be considered because stereoscopy is a very powerful means for providing a realistic spatial impression of the presented scene [1]. Stereoscopy can be viewed with special glasses or special display devices. Since portability is very important for mobile devices, special display devices attached to mobile devices are proper.

In this paper, we propose an autostereoscopic image rendering system for mobile devices. The proposed system consists of two parts: hardware configuration for the autostereoscopic display and the stereoscopic rendering software engine. In this paper, we employ a special display device which is converted from the monocular view mode to the binocular mode and vice versa. Our rendering software engine can generate both monocular-viewed images and binocular-viewed images and is tested on various types of mobile platforms such as pocket PCs, handheld PCs, and mobile phones. In the proposed system, the left-eye image is rendered only on odd lines and the right-eye image is rendered only on even lines not to render unnecessary lines of each image, so computational burdens of mobile devices are alleviated.

The rest of this paper is organized as follows. After we present the hardware configuration for rendering autostereoscopic images in Section 2, Section 3 describes

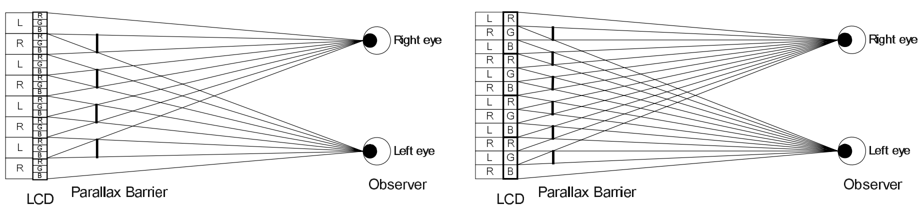
S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 441–445, 2004.

the stereoscopic rendering. In Section 4, we discuss implementation results, and concluding remarks follow in Section 5.

2 Autostereoscopic Display System

In this paper, we use the stereoscopic images, which consist of left and right images, and those images are interlaced to provide three dimensional effects using direction-selective elements such as parallax barriers or lenticular sheets [3, 4]. Parallax barriers, form of fixed film or TN panel, have a function as a spatial de-multiplexer to separate left-eye and right-eye views from a 3D scene. Therefore, in this paper, we make the autostereoscopic display system by the parallax barrier method using the TN panel to display stereoscopic images both in 2D and in 3D modes. The autostereoscopic display system does not require to wear any device to separate left-eye and right-eye views, that is, the autostereoscopic system sends those views to the corresponding eyes.

Typical emissive displays radiate lights equally in all directions. In order to create a twin-view autostereoscopic display, a half of pixels must only radiate lights in directions seen by the left eye and the rest pixels in directions seen by the right eye. The parallax barrier is the simplest way to block lights using strips of the black mask. The principle of the two-view parallax barrier is illustrated in Fig. 1. Fig. 1(a) is a stereoscopic display based on pixels and Fig 1.(b) is a stereoscopic display based on sub-pixels. The sub-pixel based stereoscopic display can reduce annoying effects from strips of the black mask. The left and right images are interlaced in columns on the display. The barrier is positioned in order that left pixels of images are blocked from the region of the right viewing windows and vice versa.



(a) Pixel-Based System (b) Sub-Pixel-Based System

Fig. 1. Autostereoscopic Display Panel Geometry using Parallax Barrier Method

The autostereoscopic display system has the capability to switch from the 2D mode to the 3D mode electronically and vice versa [5]. Fig. 2 shows the structure of the autostereoscopic display panel. The light is filtered at the first polarizer, modulated by liquid crystal, and filtered at the second polarizer. Here, the characteristic of the first polarizer is opposite to that of the second polarizer. In the 2D mode, since the TN panel does not operate, the modulated color light passes the TN panel and the third polarizer without any changes. Therefore, the autostereoscopic display system operates like conventional LCD displays. In the 3D mode, however, the autostereoscopic display system sends left-eye and right-eye images to the corresponding eyes because the TN panel operates as the direction-selective element.

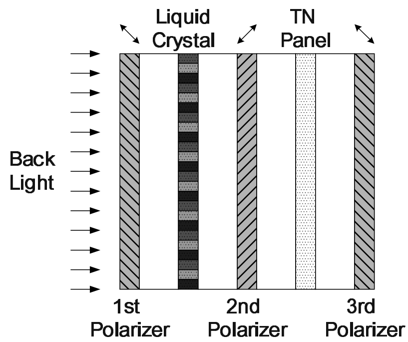


Fig. 2. Structure of Autostereoscopic Display Panel

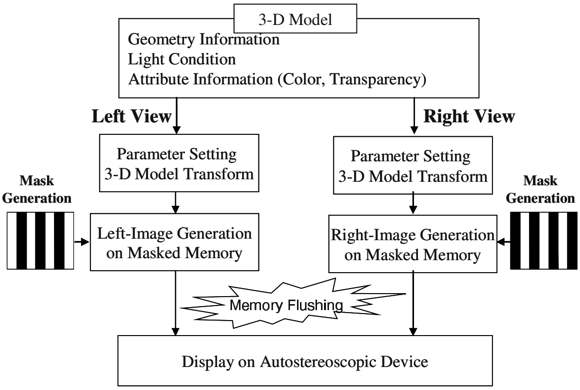
3 Stereoscopic Rendering for 3-D Graphic Models

In general, the most important factor to let human being feel three dimensional effects is spatial differences between left and right retinas. These differences are generated from slightly different view points of left and right eyes. However, conventional 3-D graphic rendering libraries set imaginary three axes on a virtual 3-D space and render 3-D graphic models on a monoscopic display device. Therefore, it is not proper for conventional 3-D graphic engines or games to provide three dimensional effects because they render graphic models only from one view point.

In this section, we propose a stereoscopic rendering algorithm for 3-D graphic model to provide better three dimensional effects to users. Fig. 3(a) shows the rendering procedure to generate autostereoscopic images on mobile devices. This procedure receives 3-D computer graphic model data as input data. These data consist of geometry information which describes the shapes of 3-D models, light conditions which represent light position and light intensity, and attribute information which describes texture information, transparency and reflection coefficients. Fig 3(b) shows a GPRS phone with the autostereoscopic display and rendering system.

For stereoscopic images, we generate left and right masks to allow left-eye and right-eye images to be displayed on odd and even lines of the autostereoscopic display device, respectively. In the proposed rendering procedure, the left-eye image is rendered only on odd lines and the right-eye image is rendered only on even lines not to render unnecessary lines of each image, so computational burdens of mobile devices are alleviated.

In order to generate left-eye and right-eye images, we set parameters to generate parallax, and perform transformation of 3-D models. Convergence distances and convergence angles from viewing points are set to generate left-eye and right-eye images. These parameters provide binocular disparities to users, so users feel sensation of reality. After setting parameters, such as convergence distances and convergence angles, we rotate the entire space that 3-D models are represented on. The entire space is rotated in the counterclockwise direction for the left-eye image, while the entire image is rotated in the clockwise direction for the right-image, because left and right eyes see the right side and left side of 3-D models, respectively.



(a)

(b)

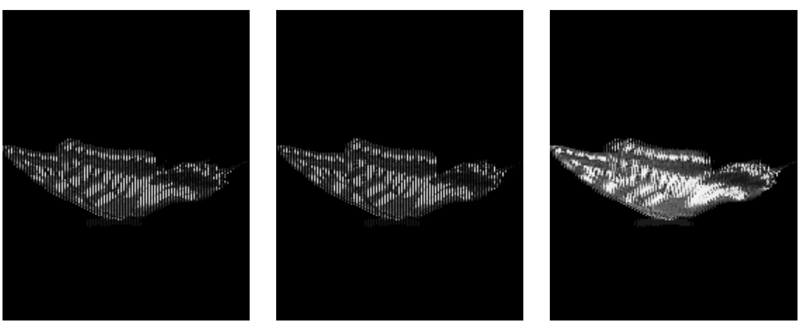
Fig. 3. Rendering Procedure for Autostereoscopic Rendering

After the rotation for each view, every 3-D model is rendered on the rotated space in conventional rendering method with light conditions, texture information, alpha blending information and etc. Since the computation capability of mobile devices is much inferior to that of general PCs, we employ fixed-point operations rather than floating-point operations, look-up tables to deal with trigonometrical functions, and shift operations rather than multiplication operation [2].

Finally, our rendering procedure displays the left-eye image in odd lines and the right-eye image in even lines using masks, simultaneously.

4 Implementation Results

In order to support various mobile service environments, we have implemented and tested our autostereoscopic rendering system on various embedded operating systems, such as Windows CE 3.x, Windows CE.Net and Nucleus. In addition, we have verified our system on the various mobile device platforms such as pocket PCs, handheld PCs and mobile phones.



(a) (b) (c)

Fig. 4. Rendering Results for Autostereoscopic View

Fig. 4 shows implementation results from the butterfly model. Fig. 4(a) is the left-eye image on odd lines and Fig 4(b) is the right-eye image on even lines. In Fig. 4, the left-eye image represents the right side of the butterfly while the right-eye image does the left side of it. Fig. 4(c) is the final output image for the autostereoscopic display devices.

As mentioned in Section 2, the full-pixel parallax barrier might generate thin black vertical lines. To address these problems, we have also employed the sub-pixel parallax barrier. In order to utilize the sub-pixel parallax barrier, the proposed rendering procedure needs post-process that exchanges green components of odd lines for green components of even lines among color components.

5 Conclusions

In this paper, we have presented a stereoscopic image rendering method for autostereoscopic mobile devices. The proposed method consists of autostereoscopic display system and the stereoscopic rendering software engine for mobile devices. In order to display stereoscopic images, we mount the parallax barrier, which is made by the TN panel, on the LCD. In addition, we have presented the stereoscopic rendering algorithm for 3-D graphic models. The proposed algorithm generates left-view images and right-view images from 3-D graphic models. Therefore, the proposed rendering system provides autostereoscopic views in order that users can enjoy three dimensional effects without any special glasses. Additionally, the proposed system is applied to produce 3-D games and 3-D contents which are competitive in mobile markets.

Acknowledgment. This work is supported by the Ministry of Information and Communication of Korea under title “The development of SmarTV technology.” In addition, the authors wish to thank Nexxpace Co. (http://www.nexxpace.com) for supporting equipment.

References

1. Runde, D.: How to Realize a Natural Image Reproduction using Stereoscopic Displays with Motion Parallax. IEEE Trans. CSVT, Vol. 10, No. 3, (2000) 376-386
2. Harbor, J.S.: Pocket PC Game Programming, Prima Publishing, Roseville (2001)
3. Woodgate, G., Harrold, J., Jacobs, M., Moseley, R., Ezra, D.: Flat Panel Autostereoscopic Displays - Characterization and Enhancement. Proceedings of the SPIE, Vol. 3957 (2000)
4. Susumu, I., Nobuji, T., Morito, I.: Technique of Stereoscopic Image Display. EP Pat., No. 0 354 851, (1990)
5. Ezra, D., Woodgate, G.J., Omar, B.A., Holliman, N.S., Harrold, J., Shapiro, L.S.: New Autostereoscopic Display System. Sharp Technical Journal, Vol. 62, (1995) 10-14

The Effects of Mobile Pedestrian Navigation Systems on the Concurrent Acquisition of Route and Survey Knowledge

Antonio Krüger, Ilhan Aslan, and Hubert Zimmer

Dept. of Computer Science, Dept. of Psychology: Brain & Cognition Group
Saarland University, Germany
{krueger,ilhan}@cs.uni-sb.de, huzimmer@mx.uni-saarland.de

Abstract. In this paper we report results of an experiment that investigates the effects of mobile pedestrian navigation systems on the development of route and survey knowledge acquired by the users. In the experiment directions were presented incrementally step-by-step in different modalities (i.e. audio, graphics) and through different media (PDA, clip-on display). The experiment has been carried out in the field in a Wizard-of-Oz like study. Results show that as expected all subjects had problems in building up survey knowledge of the environment. In contrast, route knowledge was learned much better. We also observed a slight gender effect showing that women had an advantage of a visual presentation condition, whereas for men the presentation mode didn't matter. Finally, we discuss some implications on the design of pedestrian navigation systems.

1 Introduction

On the one hand, pedestrian navigation systems seem to have the potential to provide useful mobile assistance in unknown environments. On the other hand, the ubiquitous availability of the assistance might have the side effect that users do not make an effort to acquire spatial knowledge because such knowledge is not necessary any longer for reaching the destination. From experience we do not yet know how such kinds of systems influence the acquisition of knowledge on route and environment information, and hence we ran the following experiment. We tested the spatial knowledge that was remembered after participants navigated through an unknown terrain (a zoo) guided by a pedestrian navigation system.

Usually, humans acquire spatial knowledge about landmarks and their locations within the environment. Because this is knowledge about locations within a two-dimensional coordinate system in a more global reference system, it is often called a mental map [1] or survey knowledge. Such maps can be used to predict the spatial relations between landmarks, as for example the direction in which a destination is located relative to the actual own position or another known landmark [2]. Survey knowledge is acquired from physical maps but together with landmark knowledge it is acquired also from active exploration of the environment [e.g., 3]. Our research has been motivated by our own experiences with car-navigation systems that provide incremental (i.e. step-by-step) instructions in guiding drivers to their destination. Our

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 446–450, 2004.

feeling was that those systems do not provide much survey information on the environment (i.e. the position and directions of relevant landmarks within a global frame of reference). This can lead to problems if drivers have to reorient themselves either due to technical deficits of the system (e.g. if satellites are not available) or due to dynamic changes in the environment that are not reflected in the data (e.g. blocked roads). In those situations drivers have to find their way on their own and usually need to rely on survey knowledge of the environment to find deviations or short cuts from their current position to their destination. Due to the nature of use of pedestrian navigation systems these problems will occur more often. The lack of GPS-signals on narrow streets and pedestrian zones are occurring more frequently and additional positioning systems (e.g. odometers), which are able to counterbalance these effects are usually not available. Under such circumstances it seems to be important to design pedestrian navigation systems in a way that the concurrent development of survey knowledge is continuously supported. The results reported in this paper originate in a first experiment aiming at investigating the influence of differently designed pedestrian navigation systems in this respect. In both designs we provided landmark (LM) information together with the to-be-taken directions. We used pictures of intersections and decision points – i.e. locations where walking directions were changed – in a viewer-centred perspective as landmarks [4] and we provided the direction in two different ways. In the visual condition we presented a line on each picture which indicated the trajectory of the intended path. In the oral condition we placed a read dot at the location where the direction had to be changed, and we presented the new direction by verbal means via headphones (cf. Figure 1). Because the visual, but not the oral condition provided LM and direction information in an integrated manner, we expected that the visual condition caused better memory than the verbal one [5]. Participants used one of these systems for navigation, and afterwards when they had reached the destination they were administered to an unexpected memory test for their landmarks memory and for their survey memory. Because males and females differ in spatial navigation and especially in the usage of LM [e.g., 6] we additionally introduced gender as further independent variable. Half of the participants were male and half female.

Fig. 1. An example of the picture of a landmark visible (a) in the visual version and (b) in the oral version of the task

2 Experimental Design

The experiment has been carried out in the zoo of Saarbrücken, which has a fairly complex network of small paths and routes. All 32 subjects were unfamiliar with the topology of the zoo and between 15 and 40 years old. A specific route in the zoo consisting of 15 street segments and 16 major decision points had been chosen for the ex-

periment. Decision points in this context are crossroads with unique appearances. Throughout the trial every subject had to walk along the same route and pass the decision points in the same order. At every decision point an image of the decision point was presented either on a PDA[1] (16 subjects) or a head mounted clip-on[2] display (16 subjects). The images were augmented to clarify the given directions. Directions were given either via audio (16 participants) or visually (16 participants). In the visual condition (Figure 1a) a line indicated the direction to take and in the audio condition images were augmented with a virtual reference point (Figure 1b) to clarify the directions given through the audio comment (e.g. "Turn right at next decision point"). It is known that each of the decision points has its own local reference system [7]. However, by augmenting pictures with reference points, it was possible for us to use a general reference model that allows comparing results between locations. After a short explanation, subjects had no problems in understanding the audio commentary.

The experiment consisted of two parts, a *study* part where participants were taken on a walk through the zoo for about 20 minutes, of course, without telling them in advance about the later test, and a *recall* part where participants were tested to investigate how much they remembered about the route they had taken. During the trial, subjects were told to go ahead, while the experimenter followed with a separate handheld at a distance of 5-20 meter behind them. The experimenter had the task to trigger the presentation of route instructions via a wireless LAN connection with the help of the second PDA. After the walk through the zoo, the subjects had to self-assess their own spatial abilities by completing a questionnaire, This questionnaire was introduced to find out whether individual differences in the preferred navigation behaviour interact with the presentation mode [7,8]. To test the directional knowledge (i.e. the ability to remember what direction had been taken at which decision point), the same images that were presented during the walk – without the lines – were displayed on a tablet PC in random order. The images contained sensitive areas that could be tapped by the subjects to indicate the direction that they had taken at each decision point. Before making a decision, subjects were asked to judge their confidence by pressing one of two additional buttons labelled "sure" and "unsure".

For testing survey knowledge, subjects had to position thumbnails of the decision points on an area that represented the zoo. In one version only the start position of the route was marked with a spot and no further information was given. In a second version the subjects had to place the thumbnails on a road-map of the zoo .All LMs were lined up at the left and right of the zoo map and participants moved the thumbnails via drag and drop to the position where they thought this LM is located.

3 Results

We counted the number of correct directions in the LM direction task, and the Euclidean distance between correct and selected location. Because the clip-on and the PDA presentation did not differ, we collapsed the data across these conditions. The results are given in Figure 2a and b. The data clearly confirm the hypothesis. Mobile naviga tion systems providing step-by-sep instructions caused a considerable LM knowledge but they did not help much in building up survey knowledge of an unknown environ-

---

[1] HP iPAQ 5450
[2] MicroOptical

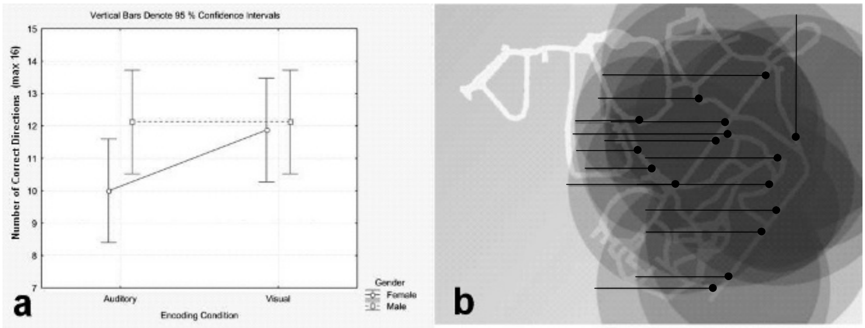


Fig. 2. Results show a gender effect for the presentation mode in the landmark direction task (a) and a poor survey knowledge (bad recall of landmark positions) (b).

ment. When looking at the results of the directional test (see Figure 2a) one may note that direction memory was relatively good. In the average 12 out of 16 directions were correctly remembered after walking the route once which is a quite good memory performance considering the incidental study condition. However, we have to take into account that not only the PDA information was available but additionally the real environment was perceived, and that participants walked without any concurrent task so that they had plenty of time to encode the surrounding. Additionally, a gender effect can be observed. Females performed worse under the audio condition (63 % correct) than the visual condition (75 % correct, $F(1,28)=3.07$, $p<.06$), whereas males always remembered 75 % correct. An explanation of this effect could be the fact that a priory females tend to remember path descriptions by verbal means [9]. Hence in the oral condition, they probably have used a verbal encoding, which is less memory efficient than a visual strategy for learning a LM-direction association [5]. In the visual presentation mode the stimulus provided environmental support for a visual encoding, and this enhanced memory. In contrast, males might always encode LMs in a visual code– even in the oral condition – so that they did not benefit from the provided visual information, or perhaps more correctly they were not harmed by oral presentation conditions. Figure 2b shows the accumulated results of the survey test. The dots indicate the correct positions of the 16 decision points. The overlapping circles represent the standard deviation of the placing (averaged over all subjects it is nearly 50% of the relevant map size). The lines represent the radius of each circle and give a quantitative idea of the differences of placing decision points. In this test the gender had no significant influence on the subject performance. On the contrary, memory was generally very poor. Obviously participants did not acquire survey knowledge during tour guided by the pedestrian navigation system. First evaluations of the questionnaires indicate that subjects' meta-memory of their own spatial abilities is internally consistent. Answers to different questions on the same topic were logically correct.

4 Design Implications for Pedestrian Navigation Systems

The first evaluation of the results has shown that mobile pedestrian navigation systems that mainly rely on step-by-step instructions are bad at conveying survey knowl-

edge of the environment to their users. This can have bad implications if the system looses the ability to determine the position on its own (e.g. when it looses GPS signal). Therefore we are currently aiming at designing pedestrian navigation systems that help to build up survey knowledge concurrently all the time throughout its use. We plan to use our general purpose navigation platform M3I [10] that can be configured in such a way to make use of two synchronized displays, using the latter one to display always a map of the environment. Another idea is to emphasize the actual decision point on the map and to visualize the spatial relationship between local landmarks on the map. Considering the individual differences and the consistencies between the personal beliefs (meta-cognition) and actual performances, it might also make sense to be able to configure such a navigation system to match the preferences of their users. This would lead to a class of navigation systems that provide spatial assistance on different levels of granularity and by different modes. The M3I platform allows for multimodal (speech & gesture-based) interaction. We believe that allowing users to interact with a map interactively and to query information about landmarks should have a positive effect on the acquisition of survey knowledge.

5 Future Work

Next steps are the investigation of the influence of different types of maps that are presented in-between the directional instructions on the acquisition of survey knowledge. Based on these future experiments we will continue to refine the design of a pedestrian navigation system that provides more than just step-by-step instructions and simple maps.

References

1. Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review,* 55, 189-208.
2. Zimmer, H. (2004). The construction of mental maps based on a fragmentary view of physical maps. Journal of Educational Psychology. (in press).
3. Aginsky, V., Harris, C., Rensink, R., & Beusmans, J. (1997). Two strategies for learning a route in a driving simulator. *Journal of Environmental Psychology*, 17, 317-331.
4. Steck, S. D., & Mallott, H. A. (2000). The role of global and local landmarks in virtual environment navigation. *Presence*, 9, 69-83.
5. Zimmer, H, Seidler, B.,. & Baus, J. (2004). Utilizing the picture superiority effect at route learning. (submitted for publication)
6. Saucier, D., Bowman, M., & Elias, L. (2003). Sex differences in the effect of articulatory or spatial dual-task interference during navigation. *Brain and Cognition,* 53, 346-350.
7. Hegarty, M., Richardson, A., Montello, D., Lovelace, K,. & Subbiah, I., (2002). Development of a self-report measure of environmental spatial ability. Intelligence, 30, 425-448.
8. Garden, S., Cornoldi, C., & Logie, R. H. (2002). Visuo-spatial working memory in navigation. *Applied Cognitive Psychology,* 16, 35-50.
9. Saucier, D., Bowman, M., & Elias, L. (2003). Sex differences in the effect of articulatory or spatial dual-task interference during navigation. *Brain and Cognition,* 53, 346-350.
10. Wasinger, R., Stahl, C., Krüger, A., (2003) M3I in a pedestrian navigation & exploration system, Mobile HCI 2003, Springer.

Supporting Museum Co-visits Using Mobile Devices

Yann Laurillau and Fabio Paternò

Consiglio Nazionale delle Ricerche (CNR-ISTI),
Via G. Moruzzi 1, 56100 Pisa, Italy
{Yann.Laurillau, Fabio.Paterno}@isti.cnr.it

Abstract. The goal of this work is to provide tools that promote social interactions between visitors through cooperative and educational games. In this paper, we describe how to support collaborative learning in museum visits and show an example application based on mobile palmtop systems. To this end, we have developed a system that is able to support collaborative and independent activities, and offer context-aware content.

1 Introduction

The wide dissemination of mobile technologies, such as cell phones or handheld personal digital assistants (PDAs), offers a good opportunity to get groupware applications out of the laboratories and develop new kinds of groupware applications which are no longer reserved for professionals and desktop computers. Mobile devices are becoming real social media, particularly in terms of communication. Such technologies are means to explore collaborative activities and move groupware applications to public settings, such as museums.

The museum visit is usually perceived as an individual experience. Furthermore, electronic guides or interactive systems in museums are not designed to promote social interaction among visitors. However, the museum experience, according to Falk and Dierking [3], is influenced by the social context, which includes interactions between visitors. In addition, many studies have highlighted the fact that interactions with the exhibit, as well as communication and social interaction between visitors are also key points of a successful learning environment [6] [7].

The research on social interaction and collaboration using new technologies is quite recent. Interest has grown with the evolution of mobile devices. In addition, there has been a change in the design of museum exhibitions in an increasing number of projects: little by little, the museum experience is considered as a collaborative activity and, more and more, museums are designed to support and encourage group interactions.

In this work, our goal is to promote interaction and communication between visitors through cooperative and interactive educational games, based on sharing, and using handheld PDAs. In this context, museum interactive systems are embedded in an electronic companion rather than being static fixtures in the museum. In addition, interactivity is considered at the visit level and not only at the artwork level: visitors are able to pace the visit and interact in the museum according to their desires. Fur

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 451–455, 2004.

thermore, educational games are an interesting and entertaining way to initiate and promote collaboration between visitors. For example, the Ghost ship project [5] shows that playing and exploring artwork may help visitors to initiate collaboration. However, in this project, the Ghost Ship is a single interactive artwork and the approach should be extended to all the artworks in the museum.

In order to obtain a new solution, we have developed and deployed an interactive system, the collaborative extension of the portable Cicero [2], dedicated to supporting the visit of the Marble Museum of Carrara. This system enables communication, sharing and collaboration among visitors, and also offers context-aware and personalized content. In the rest of the paper, we detail the main ideas of this project and provide a short review of related work. We then introduce our approach to support co-visiting in museum environments through PDAs. In the third part, we describe our system, the portable Cicero system.

2 Museum Co-visiting

Museum co-visiting has been considered in a number of projects. The Sotto Voce project [4], developed at Xerox PARC, is a mobile companion, based on the iPaq technology, that provides audio content of artwork descriptions and acts as an audio media space between visitors, which offers a mean for awareness and sociability. The authors have identified four kinds of activity: (i) *shared listening*, in order to promote interaction and communication between companions; (ii) *independent use*, in order to enable temporarily or entirely the switching off of the shared listening, in particular when visitors do not want to engage social interactions; (iii) *following*, when a companion is in charge of driving, implicitly or explicitly, the tour; (iv) *checking in*, which is a short activity, to maintain and update the shared context.

The City project [1], part of the Equator project, takes place at the lighthouse in Glasgow, a museum dedicated to the work of the designer Mackintosh. The system considers three kinds of technology: (i) a real visit using a PDA; (ii) for the virtual reality visit in a 3D world; (iii) for the Web visit. With this system, visitors are able to share their museum experience visit and navigate jointly through mixed realities: the Web, the virtual and physical reality. Information is provided about each visitor location and orientation. In addition, they may communicate through audio channels. The authors have observed that voice interaction, location and orientation awareness, and mutual visibility are essential to the success of museum co-visiting between remote users.

The Ghost Ship project [5], compared to the previous projects, is more oriented to an artistic experience of the museum co-visiting. The goal of this work is to analyse and consider informal and social interactions between visitors through video interaction recordings. The Ghost Ship installation is a dedicated room of the SOFA exhibition containing a wood painted ship, wooden figures, a simulated desk and an "inside the ship" area. Some of the ship portholes are video portholes which record and show visitor's behaviours and interactions with the ship. In addition, microphones capture visitor comments about their actions and about what they can see on the video portholes. The authors observed that the Ghost Ship helps visitors to break the ice more easily and to play with and explore collaboratively the ship.

Compared to the Equator City project, we consider "physical" visitors moving in the real museum while they consider a mixed visit combining the real museum and a virtual representation of the museum (in a 3D representation or through a Web site). Some of the existing projects consider pure collaborative virtual visits as Web co-visiting such as the Van Gogh museum. The authors of the Equator City project, as in the Sotto Voce project, have noted that information about location and orientation of the companions (*checking in* task) is essential in a cooperative visit in order to maintain group awareness. This point has been considered in our project, as detailed in the next section: visitors are able to check for their companion and are aware about the state of the cooperative game.

3 The Cicero Project

Cooperative visiting through educational games, as explained in the introduction and highlighted by the authors of the Ghost Ship project, is an interesting way to promote co-visiting and to engage visitors to share their museum experience. In addition, it may also preserve the individual aspects of the museum experience as highlighted by the authors of the Sotto Voce project. However, audio sharing, as described and implemented in Xerox's project, may lead too much to a passive collaboration between companions such as the following or checking in tasks.
In this project, our goal is to design a new interactive and multi-user system for mobile devices that is able to support and promote social interactions in museums through collaborative activities (CSCW) based on educational games. We consider the gaming approach as means to learn things about the museum in an entertaining context. However, our work is driven by HCI and CSCW issues rather than educational issues. In our system, we consider two kinds of cooperative games: one to support explicitly cooperation and sharing; and one to support implicitly cooperation through individual activities. The first kind of cooperative game is similar to enigma or treasure hunt games: visitors have to gather clues and to solve cooperatively enigmas in order to find the solution and to find a particular artwork in the museum. This game needs visitors to share and to debate about what they have seen and learned during the visit. The second kind is a collection of educational games to discover the museum, at the individual level, all along the visit and to gather clues. Indeed, solving an individual game would provide clues to solve the shared enigma (solving a puzzle and answering to a quiz) and would provide awareness information about each visitor's activity. In addition, each visitor can pace its own visit: the group interaction is not highly coupled and the system supports mixed synchronous and asynchronous modes. A scenario is provided at the end of the section, in order to illustrate both kinds of games and how explicit and implicit collaboration is supported.

During the visit, at any time, visitors can use the museum map and the peripheral information about other visitors in order to share their clues and try to solve the shared enigma. In addition, visitors are able to submit solutions, which are validated when visitors need to meet each other and to discuss about the solution. However, in order to stimulate visitors to play with our interactive electronic guide, and their name appear in the fame list in function of the number of points accumulated during the visit: if they cooperate a lot, they receive a proportional number of points.

Designing a user interface in the context of the museum visit is not an easy task because that kind of software will be used only one time by a visitor for one or two hours: the interface must be highly intuitive and affordable at first sight. It is, in one sense, a throw-away interface. For this reason, we have tried to avoid a cognitive and visual overload of the new user interface, and we added only few icons, as shown in Figure 1(a), in order to provide information about other users and the group activity, as well as about the available games. In the new interface, visitors are identified by their name and by a coloured bullet. In addition, coloured bullets indicate what items had been seen by other visitors. An icon representing a symbol indicates that an interactive game is associated with the related artwork, which had not been already solved. The following icon 15 pts, representing two little men, indicates the current score of the group. Finally, a click on the button in the command bar, at the bottom of the screen and representing two little men and a piece of jigsaw, leads to the shared enigma screen described below.

(a) (b)

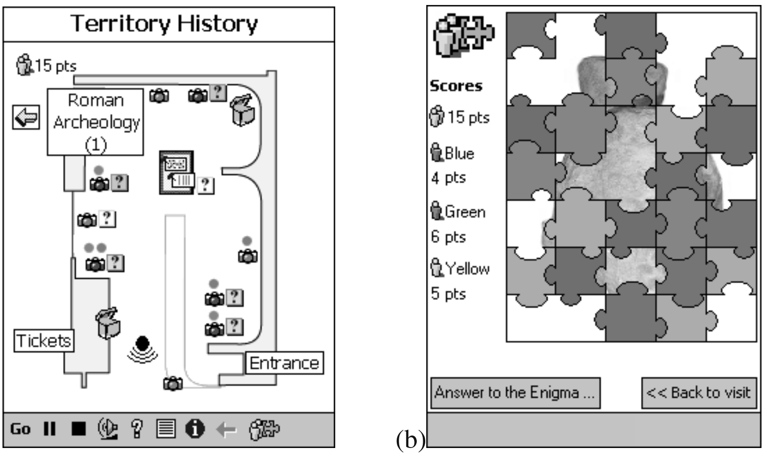


Fig. 1. (a) View of the current room (b) Puzzle part of the shared enigma.

In Figure 1(b), we present an example of shared enigma: the goal is to find which artwork is hidden by puzzle pieces and to answer to a quiz about this artwork. Each time a visitor finds a clue for the shared enigma, a puzzle piece is removed and a piece of information about the artwork is made available. On the left part of the screen, the system indicates the current score and how many points each visitors have gained. In addition, visitors are able to answer questions about this artwork even if there are still puzzle pieces hiding parts of the image. At any time, the visitors can decide together to provide and validate a common answer to solve the shared enigma, based on the set of clues gathered during the visit. To illustrate this, let us consider the following scenario: Fabio (blue player), Yann (green player) and Carmine (yellow player) are visiting the Marble Museum and have decided to play together during the visit; the goal of the shared enigma is to find the artwork representing the statue of goddess Luni and to answer some questions about it, such as “who was Luni ?” (a goddess, protecting the colony of Luni, living near the town of Carrara). During the

visit, Fabio is playing some educational games. For example, one is to associate a type of marble with the right picture; another one is to play with letters in order to find the author of Vicarius' epigraph. Fabio has solved these games and has gained two clues that are automatically shared with Yann and Carmine: two pieces of information about the statue ("it represents a goddess" and "she was the protector of a colony"); in addition, two puzzle pieces, as shown in Figure 1(b), are removed and the middle part of the statue is now visible. At that point, Yann and Carmine are aware that new clues have been found, as indicated by the "two little men" icon  15 pts, which is updated to indicate the new score, as shown in Figure 1(a). Yann has found which artwork is hidden by the puzzle using the clues discovered by Fabio, but the questions still remain unsolved. Yann asks Fabio and Carmine if they have any idea. Based on the clues, they discovered that the sculpture represents a goddess, protector the colony of Luni: the shared enigma is solved.

4 Conclusion and Future Work

In this paper, we have presented a system that enables and supports co-visiting at the Marble museum of Carrara. Compared to the existing works, the novelty of our project is to promote communication and social interactions between visitors based on interactive and cooperative educational games embedded in mobile devices such as iPaq PDAs. Furthermore, with this system, we consider the individual museum experience and visitors can regulate the pace of their own visits more flexibly than in solutions as those proposed by the Sotto Voce project where companion visitors share the audio comments. Finally, early feedback from the users has been encouraging. However, more formal tests are being currently conducted in order to improve our system.

References

1. Brown, B., MacColl, I., Chalmers, M., Galani, A., Randell, C., and Steed, A., Lessons from the Lighthouse: Collaboration in a Shared Mixed Reality System, in Proc. ACM CHI'03, 2003, p. 577-584, ACM Press.
2. Ciavarella, C., and Paternò, F., Design Criteria for Location-aware, Indoor, PDA Applications, in Proc. of Mobile HCI'03 Conference, 2003, p. 131-144, Springer Verlag.
3. Falk, J.H., and Dierking, L.D., Learning From Museums, Visitor Experiences and the Making of Meaning, 2000, 288 p., Rowman & Littlefield.
4. Grinter, R., Aoki, P., Hurst, A., Szymanski, M., Thornton, J., and Woodruff, A., Revisiting the Visit: Understanding How Technology Can Shape the Museum Visit, in Proc. ACM CSCW'02, 2002, p. 146-155, ACM Press.
5. Hindmarsh, J., Heath, C., Vom Lehn, D., and Cleverly, J., Creating Assemblies: Aboard the Ghost Ship, in Proc. CSCW'02, 2002, p. 156-165, ACM Press.
6. Hindmarsh, J., Heath, C., Vom Lehn, D., Ciolfi, L., Hall, T., and Bannon, L., Social Interactions in Museums and Galleries, in Deliverable of the Shape Project, D2.1, 2001, Chapter 2, p. 9-23, Disappearing Computer Project.
7. Leinhardt, G., and Crowley, K., Objects of Learning, Objects of Talk: Changing Minds in Museums, in S. Paris (Editors), Multiple Perspectives on Object Centered Learning, 2002, p. 301-324, Lawrence Erlbaum Associates.

Chording as a Text Entry Method in Mobile Phones

Seongil Lee and Sang Hyuk Hong

Sungkyunkwan University, School of Systems Management Engineering,
440-746 Suwon, Korea
{silee, skkie96}@skku.ac.kr

Abstract. A chording input method on mobile phones, using two thumbs simultaneously, was examined. This paper addresses advantages and disadvantages in chording applied to text input and control using keypads of mobile phones. Chording can be a good input and control method in that it can provide more options with limited number of keys, and that it can provide mobile users with extended usability and functionality. The chording input method on a mobile keypad showed comparable performances in choice reaction times and error rates with conventional single keying method, even though certain chording seemed to cause troubles in finger coordination, resulting in longer reaction times and more errors.

1 Introduction

Keyboards have been used as an important input device to computers. The increase in use of small portable electronic products such as PDAs, cellular phones, and other wearable computing devices requires some type of external input devices for convenient and error-free typing. In mobile devices, full-size QWERTY keyboards are generally not practical. Conventional keyboards do not meet the needs of portability for these small electronic devices, resulting in many alternatives in the form of flexible or folding keyboards, keypads, and stylus.

Miniature keypads became the most convenient interface for text and numeric input and for control and manipulation functions of mobile devices, providing reliable text entry for SMS and flexible control of the functions embedded in mobile phones. However, since most of the mobile phones with miniature keypads have fewer keys than letters and symbols used in any language, more than one keystroke is required for each character to be entered. With the limited number of keys, keypads also need to provide controls for ever-increasing functions of mobile devices due to today's trends of digital convergence, and its interface became quite naturally less transparent.

This paper examines advantages and disadvantages in chording methods applied to text input and control using keypads of mobile phones. Chording can be a good input and control method in that it can provide more options with limited keys, and that it can provide mobile users with extended usability and functionality. Just like users of conventional keyboards where they apply chording while typing "*Ctrl* + " or "*Shift* + " keys simultaneously for various functions, users in mobile devices can naturally

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 456–460, 2004.

develop effortless strategies for text entry and control tasks using the similar chording mechanism.

2 Related Studies

For use of many small portable electronic products, chord keyboards have long been proposed as input devices [1-6]. A chord keyboard is a keyboard that takes simultaneous multiple key pressings at a time to form a character in the same way that a chord is made on a piano. In chord keyboards, the user presses multiple key combinations, mainly two-letter combinations, to enter an input instead of using one key for each character. Pressing combinations of keys in this way is called chording [3, 5]. Since chord keyboards require only a small number of keys, they do not need large space, nor the many keys of regular keyboards such as the QWERTY keyboard. For example, the Handkey Twiddler is a one-handed chord keyboard with only 12 keys for fingertips and a ring of control keys under the thumb, and the Microwriter with only 6 keys. With a typical two-handed chord keyboard, most Braille writers have a keyboard of only six keys and a space bar for all the Braille characters. These keys can be pushed one at a time or together at the same time to form Braille symbols for visually impaired people [6].

Similar input mechanisms utilizing chording could be found in glove-based input devices. Pinch Gloves [1] are glove-based input devices designed for use in virtual environments, mainly for 3D navigation. Pinching is basically the motion of making a contact between the tip of thumb and a fingertip of the same hand. Rosenberg and Slater [5] proposed a glove-based chording input device called the chording glove to combine the portability of a contact glove with the benefits of a chord keyboard. In their chording glove, the keys of a chord keyboard were mounted on the fingers of a glove and the characters were associated with all the chords, following a *keymap*. For extensive review on chord keyboards, see Noyes [3].

3 Chording in Keypads

3.1 Text Entry

Keying using chording can provide faster text entry than conventional, serial, one-keystroke-at-a-time keying method. Since most mobile phones with miniature keypads have fewer keys than letters and symbols used in any language, more than one keystroke is required for each character to be entered. Many mobile phones assign two to three Roman-alphabets and one to three Korean characters onto a key (Fig. 1). Therefore, serial input requires one to three times many keystrokes as constructing letters for a word.

Chording, on the contrary, can save keystrokes by pressing two or more keys simultaneously. For example, while current commercial text entry interface for Korean

letters in Fig.1 requires 24 keystrokes for “Hello” in Korean, a new interface accommodating chording text input only requires 13 keystrokes, with a chord counted as one stroke. Chording can be particularly useful and effective for typing complex letters such as diphthongs in Korean, and common endings of a word such as “-tion” or “-sion” in English with chording two letters in such way as “t+n” or “s+n”.

Fig. 1. Keypads of Samsung’s various models of mobile phones. The number of keys on the keypad increases in time from 18 (left, manufactured in 1998), 21 (center, in 2000, with additional arrow keys), to 23 (right, in 2003, with keys for wireless internet and digital camera).

3.2 Control and Navigation

Today’s mobile phones include many functions by convergence with other digital devices. The increase in numbers of additional functions and complicated information architecture require more keys for control. Fig. 1 shows the increase in number of keys on the keypads of mobile phones. The increase in functions such as voice recording, digital camera, and wireless internet services, which services and functions had not been offered by the earlier models, required more keys for execution, control, and navigation.

Chording can let users easily execute and control special functions and services on today’s complicated mobile phones. The so-called “soft keys” on the screen that map the functions and controls on the screen to specified keys on the keypad change according to functional modes, generating cognitive mapping problems. Internet services offered in mobile phones particularly adopt use of “soft keys” for execution and navigation. This takes many steps and as many keystrokes in addition to cognitive demands in attention, since the soft keys continuously changes in each step or mode. Chording reduces the mapping by constructing “shortcuts” easily. Users can customize the frequently used functions by chording keys to construct more user-friendly control and navigation structure.

4 Chording Input Experiment

Experiments were performed to measure the input speed and accuracy for chording. The choice reaction time and resultant error rate were measured using a keypad on Samsung's mobile phone model SCH-E140.

4.1 Experiment

Subject. Eight college students (four males and four females) participated in the chording input experiment. All the subjects had previous experiences in use of mobile keypad for sending and storing short text messages (average use of 4.5 times per day) in Korean. The average age of the subjects was 23.2 years old. All the subjects were right-handed.

Procedure. Subjects' choice reaction times on the 12 keys (1 to 0, plus * and # keys) on the keypad were measured in the first session. Subjects were asked to key the same keys on the keypad with the numbers or symbols randomly shown on a computer screen with a beep as quickly and correctly as possible. The next signal came with a 5 second interval as soon as the previous trial was completed with a correct keying. A total of 10 trials for each key were performed.

In the second session, Subjects were asked to key in the required chording combination of the keys. Only four keys (1, 2, 3, and *) to make two-key chords were used in the experiment. A combination of two keys, randomly generated for 6 combinations, was shown on a computer screen with a beep, and subjects keyed in the appropriate two keys simultaneously as quickly and correctly as possible. There existed time differences between actual keying of the two keys. The reaction time to the later keying was recorded if the time difference between the two keys did not exceed 100 msec for each trial. A total of 10 trials for each chord were measured.

Table 1. Keying performances (mean and standard deviation) for each chording combination.

Chording Combinations	Choice Reaction Time (msec)	Error Rate (%)	Statistical Significance
1 Key, overall	**827.02 ± 53.64	1.4 ± 1.5	$p < 0.01$
1 + 2	831.96 ± 64.06	**0.3 ± 0.5	$p < 0.01$
1 + 3	840.73 ± 82.04	1.8 ± 1.4	
2 + 3	887.55 ± 130.85	1.0 ± 1.2	
1 + *	*933.15 ± 114.02	*2.8 ± 2.1	$p < 0.05$
2 + *	*940.53 ± 123.39	1.8 ± 1.4	$p < 0.05$
3 + *	*929.93 ± 112.37	2.1 ± 2.2	$p < 0.05$
2 Keys, overall	893.97 ± 104.46	1.6 ± 1.5	

4.2 Result

All the subjects pressed the two keys for generating the required chords with the thumbs of both hands. The results for the experiment are summarized in Table 1. Chord keying requires more time than single keying (p <0.01), and certain combinations of keys for generating a particular chord resulted in more errors than other combinations. The chording that involves the "*" key resulted in longer reaction times and more errors than other chords, suggesting difficult finger coordination for simultaneous keying motions. The chords involving the same side but different rows such as "* + 1" combination and possibly "* + 2" combination caused troubles for many subjects in finger coordination. Most subjects chose the left thumb for pressing the "*" key and the right thumb for the number key.

5 Discussion

The chording has distinct advantages over conventional single keying method in miniature keypads in that chording can let users easily execute and control special functions and services on today's complicated mobile phones. Even though two-finger chording requires more time to execute and results in higher error rates than single keying, it provides more options in information architecture and navigation strategies in constructing user interfaces of mobile phones.

Two-finger keying performances for generating chords can be enhanced with more training once users are acquainted with the method. Future studies need to focus on effective strategies on keymap design for chording and required memory burden accompanying the chording. The performances on real text entry tasks measuring keying speed in terms of words per minute and accuracy also need to be examined for evaluating its usefulness.

References

1. Bowman, D. A., Wingrave, C. A., Campbell, J. M., Ly, V. Q.: Using Pinch Gloves for both Natural and Abstract Interaction Techniques in Virtual Environments. In: Proceedings of the HCI International 2001, Lawrence Erlbaum Associates, (2001) 629-633
2. Gopher, D., Raij, D.: Typing With a Two-Hand Chord Keyboard: Will the QWERTY Become Obsolete? IEEE Trans. On Systems, Man, and Cybernetics, 18(4) (1988) 601–609
3. Noyes, J.: Chord keyboards. Applied Ergonomics, 14(1) (1983) 55-59
4. Porosnak, K. M.: Keys and Keyboards. In: Helander, M. (ed.): Handbook of Human-Computer Interaction, Elsevier, New York (1988) 475-494
5. Rosenberg, R., Slater, M.: The Chording Glove: A Glove-Based Text Input Device. IEEE Trans. On Systems, Man, and Cybernetics-Part C: Applications and Reviews, 29(2) (1999) 186-191
6. Weber, G.: Reading and pointing-New interaction methods for Braille displays. In: Edwards, A. (ed.): Extra-Ordinary Human-Computer Interaction, Cambridge University Press, Cambridge (1995) 183-200

Engaging Learners with Everyday Technology: A Participatory Simulation Using Mobile Phones

Peter Lonsdale, Chris Baber, and Mike Sharples

University of Birmingham
Birmingham B15 2TT, UK
p.lonsdale@bham.ac.uk

Abstract. Mobile text messaging provides a novel opportunity to engage learners in collaborative learning experiences. This study describes the use of mobile text messaging to engage learners in an interactive role-playing game based on the water cycle. Learners were guided through their role in this participatory simulation using SMS messages sent to their mobile phones. Learners' physical positions in the game triggered particular events and precluded others, thus enabling an interactive simulation that was more than just a series of messages sent out at fixed intervals. Results indicated that the learners' understanding of the topic was at least as good as in a comparative conventional teaching condition, despite being given far less causal information. Furthermore, a set of learners who answered a test of water cycle knowledge without any teaching gave different responses to the SMS group, suggesting that the SMS game is at least a viable and entertaining alternative to lesson-based teaching for appropriate topics.

1 Introduction

There is a growing level of interest in how we might use games to provide engaging and effective learning experiences. One focus of current attention is predominantly the educational use of computer or video games, but gaming is not restricted to play with or through the device. Interactive games can take the form of wider, distributed games that involve both face-to-face interaction and mobile devices such as PDAs and phones. Games where players take an active role and act out the game themselves in a real, physical environment are perhaps even more compelling and effective than games limited to the screen of a computer. For example, a game where learners are actually out in the field, following a pre-defined scenario, with mobile devices performing maintenance of that scenario, could be more compelling than learners playing a simulation at a desktop PC. The use of mobile devices to enable these kinds of interactions can be found in recent work such as the Ambient Wood project [5] and Savannah [3]. The current research aims to address the need for further study of the use of mobile devices to engage learners in *participatory simulations* (see [6]).

Participatory simulations are learning games where players play an active role in the simulation of a system or process. Simulations of this type have recently come to

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 461–465, 2004.

the attention of educators through the work of Colella [1]. In this study, participants used bespoke 'Thinking Tags' to observe the spread of a disease in a population. Learners became actively involved in the simulation, and were able to test out their hypotheses about the spread of the disease.

The proliferation of mobile devices such as PDAs means that classrooms will increasingly have the means to run games of this type. However, these technologies are still relatively expensive, and require specific programming and configuration. There is another type of device that has already found its way into the classroom, which could also be used to play these kinds of games: the mobile phone. Mobile phones offer a novel, engaging, and easy way to engage learners in interactive learning games. Although the retrieval and display capabilities of typical phones are limited, we can still use them as effective facilitators for learning by embedding them in the larger system of a participatory simulation.

The display capabilities of contemporary phones surpass those of the original Thinking Tags used in Colella's study, and offer the means to engage the learner in a personal way. Whilst text messages may potentially be viewed as a private and personal means of communication, it is possible to harness this medium to enable people to communicate information relating to a specific topic via a mediating host [7]. It is hoped that utilizing this medium to facilitate learning will engage learners in a positive and personally engaging way. Although SMS has been used before now to engage learners in learning activities (such as question & answer, reminders, and informal chat) it has not been used to engage them in active and participatory learning. The advantages of learning in this way include:

- engagement – learners are drawn in to the learning experience by having to play a role, react to instructions, and process information;
- collaboration – playing the game means playing with other people, and finding out what they think about what is going on. Different game states are represented not on the screen by the actions of others, clearly visible in the physical world;
- Learning through doing – The abstract becomes concrete: concepts that are not easily represented in text or even animations become clearer as players act out their roles, and see the causal relationships between themselves and other players;
- real-world interaction – the learning experience involves interacting with the real-world, thus grounding the experience for the learner;
- context awareness – the state of the game system is affected by the states of the players, meaning that the game is sensitive to context and players are able to observe changes in the game due to their actions and the actions of others.

This study presents the use of mobile phones to enable a participatory simulation of the water cycle, sending out information and instructions to players via text messages, and tracking their location on a physical representation of the water cycle by means of real-time observation.

2 Implementation

The topic chosen for this study was the watercycle. This is a set of relatively simple natural processes that school children learn about in the classroom. This topic was felt to be appropriate due to its cyclical nature. A simplified version of the water cycle was used for this study, focusing on evaporation of water from the ocean, condensation of water vapour into clouds, precipitation of water as rain, and surface run-off of water back into the ocean.

All participants were undergraduate students at the University of Birmingham's Department of Electrical, Electronic, and Computer Engineering. While the scenario is ultimately aimed at school-children, it was felt that these participants were appropriate for the proof-of-concept test that was performed at this stage of the development cycle. Participants were recruited to one of three conditions:

Game condition: participants took part in the participatory simulation in groups of six (two groups, total n=12). The procedure for this condition is described in more detail below.

Workshop condition: participants took part in a classroom based session (two groups, total n=17). A facilitator led the participants in a discussion of textual description of the water cycle, with reference to diagrams drawn on a white board as part of the session.

Control condition: participants submitted individual responses via a web-based survey (n=17) without being given any information pertaining to the water cycle.

For the game condition, all participants were asked to bring their own mobile phone with them to take part in the study. The game simulation ran under Java on a Windows laptop PC. Text messages were sent via a GSM mobile phone connected by data cable. The jSMSEngine library [2] was used to implement the communication between the Java simulation and the mobile phone.

Learners played the role of water droplets in the water cycle and received information and instructions via text messages delivered to their mobile phone. The experimenter also acted as a facilitator for the game, observing the participants and providing cues when appropriate. Learners played the game on a floor mat (approximately 2m x 4m) showing the major physical features of the water cycle, i.e. the ocean, sky, and land. Players were instructed that these locations corresponded to the locations in the simulation being run on the laptop.

The simulation of the water cycle (as detailed in [4]) was implemented as a set of 'location' objects in Java, each of which had an *entry message* that players were sent when they moved into that location, an *exit message* that players were sent when they moved out of that location, and an *exit rule* that specified when players should be sent an exit message. Exit messages were either sent at a specified rate (to one player per *n* cycles) or when the number of players in a certain location reached a specified trigger level (to all players at that location). For example, exit messages for 'ocean' were sent every 5 cycles (seconds), stating "The sun is very hot today - u get warmer & warmer until u EVAPORATE up into the sky. U are now water vapour, lighter than air" – a player receiving this message was then expected to move from the 'ocean' location

Table 1. Post-task responses from all conditions

Condition	Response	Rain n	Rain %	Cloud n	Cloud %	Sea n	Sea %	Land n	Land %
Game	less	2	16.67	2	16.67	5	41.67	5	41.67
	same	0	0	2	16.67	0	0	4	33.33
	more	10	83.33	8	66.67	7	58.33	3	25.00
Workshop	less	6	35.29	3	17.65	11	64.71	14	82.35
	same	1	5.88	4	23.53	0	0	0	0
	more	10	58.82	10	58.82	6	35.29	3	17.65
Control	less	3	17.65	1	5.88	8	47.06	11	64.71
	same	4	23.53	5	29.41	1	5.88	1	5.88
	more	10	58.82	11	64.71	8	47.06	5	29.41

up into the 'clouds over ocean' location. During each cycle, each location checked to see if its trigger values had been reached and if any exit messages needed to be sent.

Movement of the players in the simulation was tracked manually by the experimenter by moving representations of each user on a graphical interface. The experimenter could also monitor the state of each location and see which messages were being sent out.

All participants were asked to complete a post-task questionnaire asking them what they thought would happen to levels of rainfall, cloud cover, sea water storage and land water storage in the event that the climate became warmer. Responses were restricted for each item to 'Less', 'Stay the same', and 'More'.

The most important item on the questionnaire was rainfall – according to the model presented in the workshop and game conditions, the amount of rainfall would increase if the climate became warmer. This answer is counter-intuitive (warmer weather suggests less rainfall, but according to the model warmer weather leads to increased evaporation and hence more precipitation) and so differences in responses to this item were expected to vary between conditions.

3 Results

The results from the post-task questionnaire are presented in Table 1. Learning outcomes varied according to condition. Participants in the game condition gave different responses. More learners in the game condition gave the answer "more" to the question about how much rainfall there would be if the climate was warmer. This answer is the one best supported by the model presented to the participants, and the higher number of 'correct' responses from the game condition suggest that there were

gains from experiencing the water cycle through the participatory simulation. In both the control and workshop conditions, responses to the rainfall question are more spread over the three possible answers.

4 Discussion

This study demonstrated a novel use of SMS text messages for the purposes of engaging learners in a participatory learning experience. The results suggest that the game players successfully learned the rules of the water cycle from playing the game, and comparison to the control condition indicates that their success was not due to prior knowledge.

Specific topics to be addressed in subsequent studies include:

- Augmented powers for role-playing: the use of mobile devices can allow players to make use of abilities they do not normally have
- Pedagogical support – mobile learning games should not used be isolation, their role is as part of a wider system of education
- Role of the facilitator – there is still a place a facilitator and guide
- Physical mediation of the game through interactions with the world: games could exploit the rich set of interactions that would be enabled by having the game respond to changes in the environment and vice versa.

The present study is preliminary work for a PhD focusing on the use of mobile technologies to enable new ways of learning through wide-area games. The potential for mobile devices to enable new and engaging forms of learning has only just begun to be tapped. Everyday technologies can enable new and exciting forms of learning by tapping into learners' imaginations and letting them play together.

References

1. Colella, V. (2000). Participatory simulations: building collaborative understanding through immersive dynamic modeling. *Journal of the Learning Sciences*, **9**(4): 471-500.
2. Delenikas, T. (2004). *jSMSEngine*, available at http://jsmsengine.sourceforge.net/.
3. Facer, K., Stanton, D., Joiner, R., Reid, J., Hull, R., and Kirk, D. (in preparation). Savannah: a mobile gaming experience to support the development of children's understanding of animal behaviour. To appear in *Journal of Computer Assisted Learning*.
4. Gaddis, C., Kurose, A., and Tillotson, J. (2004). *Water: a never ending story*, http://www-k12.atmos.washington.edu/k12/pilot/water_cycle/index.html, accessed May 2004.
5. Price, S., Rogers, Y., Scaife, M., Stanton, D., and Neale, H. (2003). Using 'tangibles' to promote novel forms of playful learning. *Interacting with Computers*, **15**(2): 169-185.
6. Roschelle, J. (2003). Unlocking the learning value of wireless mobile devices. *Journal of Computer Assisted Learning*, **19**(3): 260-272.
7. Sillence, E. and Baber, C. (2004). Integrated digital communities: combining web-based interaction with text messaging to develop a system for encouraging group communication and competition. *Interacting with Computers*, **16**(1): 93-113.

One-Push Sharing: Facilitating Picture Sharing from Camera Phones

Gary Look, Robert Laddaga, and Howard Shrobe

MIT Computer Science and Artificial Intelligence Laboratory
77 Massachusetts Avenue 32-224
Cambridge, MA 02139
{garyl,rladdaga,hes}@csail.mit.edu

Abstract. We present a "one-push" sharing feature for camera phones that allows camera phone owners to easily email pictures from their camera phone. Our approach compares favorably to that of camera phones today, which require users to navigate through a tedious series of menus. This work was the result of examining current camera phone use and designing a feature that facilitates the ways camera phones are used. The design was refined by feedback in an iterative design process. We discuss the rationale behind this feature, describe a camera application we implemented with this one-push sharing feature, and present some of the feedback users gave us during the iterative design process.

1 Introduction

The popularity of camera phones is increasing rapidly. A total of 84 million camera phones were sold worldwide in 2003, one-sixth of all mobile phones sold. This is almost a five time increase compared to the 18 million camera phones sold worldwide in 2002 [9]. Given their popularity, camera phones and their use are becoming important areas of mobile HCI research. Our research extends previous research on camera phone use by providing a customizable menu feature for camera phones. Such a menu makes it easier for camera phone owners to share pictures with others and to review their pictures at a later time.

Since most people carry their mobile phone with them, camera phones have made it much easier for people to take and share pictures from everyday life with others. As different studies have shown, capturing images of fun and spontaneous moments for the express purpose of sharing with others is one of the primary uses of a camera phone [4,5,8,10]. The emphasis on sharing pictures taken with a camera phone is also seen in the growing number of software tools (FoneBlog, BlogPlanet, KABLOG, etc) available to create "moblogs" (mobile blogs), online journals that feature pictures taken with camera phones.

Given this practice of using camera phones to capture and share everyday images with others, we examined the current interfaces that camera phones provide for sharing images via email or multi-media messaging service (MMS). This work furthers a result from Counts' study on photo sharing. Although Counts focused on how a broadcast-style sharing mechanism can be used to enhance group

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 466–470, 2004.

awareness, in his study, participants almost unanimously requested a single-click mechanism to share pictures with a smaller, select group of people [4]. Using an iterative design process [1], we developed a customizable menu feature that provides people the ability to do this "one-push sharing" of pictures they take with their camera phone. One-push sharing provides an interface that allows people to email pictures by selecting a name from a single menu, instead of navigating through a tedious series of menus as is the case today.

2 Design and Implementation

2.1 Design Criteria

Before we began implementing the one-push sharing feature, we used a semi-structured interview approach to interview 10 people who owned and used camera phones. In this group of 10 people, there were 5 different models of camera phones. The people we interviewed had owned their camera phones for a period of time ranging from 3 months to 1 year. Our interviews were focused on learning why they took the pictures they did, with an emphasis on what they did with the pictures once they took them, and what they thought of their picture management process. We were aware that many people took ad hoc pictures for themselves or to share with others, but we were also interested in what other reasons people had for taking pictures [2]. We learned the following:

– People rarely deleted pictures from their camera phones, even if these pictures were previously transferred off the phone.
– Picture taking behavior varied greatly. At the extremes, there were people who took many pictures (on average, more than 7 pictures per month), while others rarely ever used their phone's camera feature (about 2 pictures per month). Those who took more pictures shared many of them. Those who took few pictures didn't intend to share them.
– Those who did email pictures to others had a small set of people (between three and seven) to whom they sent pictures.
– All the phones had multiple options for transferring pictures to a computer (some combination of Bluetooth, infrared, email, memory card, upload to a server, or direct cable connection). Not all of these options were used because not all the camera phone owners had the necessary hardware support (e.g., the owner may not have a Bluetooth transceiver). All phones had the ability to send pictures as an email attachment and via MMS.
– All the phones used essentially the same interface to email or MMS a picture. This involved the following sequence of steps: select picture to send, choose sending method (email, MMS, IR, etc), choose "add a recipient," identify the recipient, and then send the email or MMS. This process required a user to navigate multiple menus and press numerous buttons to send his picture.

Our original plan was to create a hot key for emailing pictures to the camera phone's owner. Our first prototype had this feature, which was intended to make

it easier for the user to review and be reminded of the pictures he took with his camera phone. However, the results of our interviews and the feedback we received from users during the iterative design process led us to expand this initial goal. Given the importance of sharing pictures with a relatively small circle of people, we decided to add a customizable menu that provides a person the ability to do one-push sharing with three people chosen by the phone's owner. We chose three people because we did not want a person to have to scroll through a long list of names to do one-push sharing.

As an alternate implementation of one-push sharing, we considered caching recent email addresses to which a person sent pictures, and presenting these addresses as one-push sharing options. We decided against this option when we considered the confusion and extra effort on the user's part that would result when an infrequently used email address would evict a frequently used email address from the cache. If the list of commonly-used addresses changed, this would require a person to spend additional time scanning the list to find the address and then even more time to input the desired address when it was missing. This, in turn, may cause further disruptions, depending on the caching strategy and how the email addresses in the cache are displayed to the user.

2.2 System Architecture

For the purposes of our iterative design process, we chose to implement our own camera application using J2ME, the Java 2 Micro Edition (specifically, MIDP 2.0 and the Java Mobile Media API). Compared to writing Symbian code in C, J2ME is much easier to use in rapid prototyping, especially given the numerous code samples available from Sun and Nokia, and the fact that Java's garbage collection mechanism protects against memory leaks. Protecting against memory leaks was a significant concern of ours since not only do mobile devices have a limited amount of memory, but their memory is designed to be persistent across powerdowns, making memory leaks a severe threat to the performance and ease of redeployment of an application on a particular phone.

Our camera application is similar in functionality to the standard camera application found on all camera phones. It allows a person to take, review, and email pictures. One limitation to our application is that email is the only means for transferring pictures off the phone. However, our one-push sharing feature allows the user to specify three people with whom to share pictures. These names appear on the top-level menu in the photo gallery, thus allowing the phone's owner to share pictures with his most common recipients by navigating only one menu, instead of the four or five menus on camera phones currently available on the market. Of course, the user could also email the picture to a single address without using the one-push sharing feature (see Figure 1).

Once picture and recipient are chosen, our camera application sends this data to a server via GPRS. The server then sends the picture as an email attachment to the intended recipient. The server is a software agent written using the Java Mail API and a distributed multi-agent system called Metaglue [3,7].

Fig. 1. After taking a picture (left), selecting "Options" will bring up a customizable menu with one-push sharing options (right). A person specifies one-push sharing recipients by selecting the "One-Push settings" menu option. For presentation purposes, this figure was created using the Nokia Series 60 MIDP emulator.

3 Evaluation

A formal study of how well one-push sharing is received by camera phone users is under way. From user evaluations of our early design iterations, we have anecdotal evidence that one-push sharing is a useful feature that facilitates the common task of sharing and managing pictures people take with their camera phones.

We gathered this evidence by giving each member of a small group of volunteers (3 undergraduates and a working professional) a Nokia 3650 camera phone for a week. We told them to treat the phone as if they were considering buying it, and were given a one week trial period in which to test it out. This group was told that if they wanted to share pictures, to do it via email only. We also gave each member of a second group of volunteers (an undergraduate, 2 graduate students, and a working professional), a Nokia 6600 loaded with our one-push sharing camera application. We also told this second group to try out the phone for a one-week trial period but to use our camera application instead of the one that comes with the phone. When asked at the end of the week to rate the ease of emailing a picture to someone, the group using our one-push sharing camera application gave higher ratings than those using the standard camera application (an average of 4.5 compared to 2.9, on a scale of 1 to 5, with 5 being the highest, or easiest). In particular, one volunteer who used the standard camera application singled out as an area for improvement the numerous menus she had to navigate in order to email a picture. This feedback from our iterative design process along with previous research showing the emphasis on sharing pictures with camera phones [4,5,8,10] supports our claim regarding the value of one-push sharing.

4 Conclusion and Future Work

Taking pictures for the express purpose of sharing with others is one of the most common uses of camera phones. However, current camera phone interfaces are not designed to facilitate this process. They require a user to navigate through many menus to carry out a common task. This required navigation violates one of the common practices in user-interface design: make the common case fast, and the not-so-common case possible. To address this issue, we have developed one-push sharing, a feature that makes it easier to share pictures taken with a camera phone. This feature takes into account a person's tendency to share the majority of his pictures with a small group of others. It compares favorably to current interfaces because it allows a user to avoid a tedious series of menus in the common case. We are currently studying how this feature is being used and received by camera phone users. Among the things we are looking at is how one-push sharing to groups of people and one-push sharing via Bluetooth would be used. We are also interested in seeing if easily-taken pictures can be used to create reminders for input to intelligent reminder applications [2,6].

References

1. Barry W. Boehm. A spiral model of software development and enhancement. *IEEE Computer*, 21(5):61–72, May 1988.
2. Barry A. T. Brown, Abigail J. Sellen, and Kenton P. O'Hara. A diary study of information capture in working life. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 438–445. ACM Press, 2000.
3. Michael H. Coen. Design principles for intelligent environments. In *Fifteenth National Conference on Artificial Intelligence (AAAI'98)*, number SS-98-92, pages 547–554, Madison, WI, 1998. AAAI, AAAI Press.
4. Scott Counts and Eric Fellheimer. Supporting social presence through lightweight photo sharing on and off the desktop. In *Proceedings of the 2004 conference on Human factors in computing systems*, pages 599–606. ACM Press, 2004.
5. Okabe Daisuke and Mizuko Ito. Camera phones changing the definition of picture-worthy. *Japan Media Review*, August 29 2003. http://ojr.org/japan/wireless/1062208524.php.
6. Anind K. Dey and Gregory D. Abowd. Cybreminder: A context-aware system for supporting reminders. In *Proceedings of the 2nd international symposium on Handheld and Ubiquitous Computing*, pages 172–186. Springer-Verlag, 2000.
7. Nicholas Hanssens, Ajay Kulkarni, Rattapoom Tuchinda, and Tyler Horton. Building agent-based intelligent workspaces. In *ABA Conference Proceedings*, June 2002.
8. Anu Mäkelä, Verena Giller, Manfred Tscheligi, and Reinhard Sefelin. Joking, storytelling, artsharing, expressing affection: a field trial of how children and their social network communicate with digital images in leisure time. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 548–555. ACM Press, 2000.
9. Reuters News Service. Camera phone sales skyrocketed in 2003. *Reuters News Service*, March 29 2004. http://msnbc.msn.com/id/4623824/.
10. Anita Wilhelm, Yuri Takhteyev, Risto Sarvas, Nancy Van House, and Marc Davis. Photo annotation on a camera phone. In *Extended abstracts of the 2004 conference on Human factors and computing systems*, pages 1403–1406. ACM Press, 2004.

Delivering Personalized Local Services

László Márai[1], Javier Lafuente[1], Gábor Márton[1], Timo Perälä[2], and Mikko Tarkiainen[1]

[1] Nokia Research Center,
{laszlo.marai, gabor.marton, javier.lafuente, mikko.t.tarkiainen}@nokia.com
[2] Nokia Ventures Organization,
timo.k.perala@nokia.com

Abstract. Aspects of delivering personalized local services are briefly investigated, such as content delivery, authentication, attribute exchange, payment and user privacy. The main emphasis is on services that reach the users from their physical vicinity via their mobile handsets. The proposed solutions are demonstrated by showcasing a local service at an imaginary music store, with two alternatives of content presentation: XHTML browsing and Java MIDlet.

1 Introduction

More and more mobile handsets are equipped with short–range radio interfaces such as Bluetooth or WLAN, thus making the delivery of services to the users from their physical vicinity possible. The range of possible types of services is wide, including advertisements in shopping malls, personal assistants in shops, information kiosks on the street, to mention just a few. However, it is a long way from the basic data connectivity stage to our goal, which is to provide seamlessly working, enjoyable, non–disturbing, personalized and trusted services.

2 Main Service Aspects

Content delivery. Using *(X)HTML* pages is the most portable solution, as long as there is a way to connect to the content server via HTTP. Another possibility is delivering *Java MIDlets* to the terminal. This allows for better tailoring of the user interface to the service than with a browser–based solution, especially with non–browsing services. Delivering a *native application* is also an option, although it is the least portable solution. *Auto–launch*, i.e. starting the browser, the MIDlet or the native application when the service is available, is an important aspect in all cases.

Authentication of the user's identity can be important (1) for the user, when access to private information is in question, and/or (2) for the service, when it offers something specifically for *that person* (e.g. loyalty points for repeated visits). *Single sign–on* or *seamless sign–on* solutions, just like Passport [1] or Liberty [2], are highly desirable.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 471–475, 2004.

Personalized services are based on *user attributes*, also known as *user profiles*. Some of the attributes may already be present at the service (tied to the user's local account), but the rest is to be queried, either from the terminal device itself or from a third party. The most comprehensive set of specifications for attribute exchange to date is that of the *Liberty Alliance* [2].

Comfortable and secure *payment solutions* are desirable for local services. The use of *vouchers* or *checks* is one possibility. The user accepts (digitally signs) the check issued by the service. The check is then forwarded to the user's "bank"—which could be e.g. the mobile operator—that will eventually charge the user. Another possibility is *e–cash*, although there is still no widely accepted e–cash system in use today (except perhaps for PayPal). In most cases, these solutions impose a privacy threat on the user because of the traceability of most digital signature systems. There are also anonymous e–cash systems (e.g. Brands' e–cash) providing maximal privacy.

Privacy protection is expected to become more and more important with the increasing number of local services. *Unsolicited connections* are to be filtered out by means of personalizable filters. For the disclosure of personal information (cf. attribute exchange), the *principle of minimum disclosure* suggests that no more information be disclosed than what is absolutely necessary.

3 Case Studies

The *user story* is briefly the following. Jimbo's store is an imaginary music store where—in addition to browsing and purchasing real discs—customers can use their mobile terminals to access the local services provided by Jimbo. When a customer enters the shop, the local service is automatically offered to her mobile terminal. The service includes an online disc catalog with personalized recommendation and a jukebox playing video clips. The users can bid for video clips using their loyalty points, and the jukebox plays the clip with the highest bid at any time.

We have built two alternative implementations. The user story, as well as the terminal device (Nokia 3650), is the same in both cases. The difference is in the selected content delivery method.

3.1 XHTML over Bluetooth

Content is delivered as XHTML pages (with the look–and–feel encoded in a separate style sheet), see Figures 1 and 2. The XHTML pages are retrieved by the browser over a Bluetooth IP connection. *Browser auto–launch* is provided by a so called "HotSpotter" application.

Authentication is done using the Liberty ID-FF protocol [2]. We chose to implement the Identity Provider (IdP) on terminal so that we don't require the user to have an on–line connection.

The screenshot on the left side of Fig. 3 illustrates how the *service profile* corresponding to Jimbo's service defines the *access card* to be used upon receiving

Fig. 1. XHTML welcome page

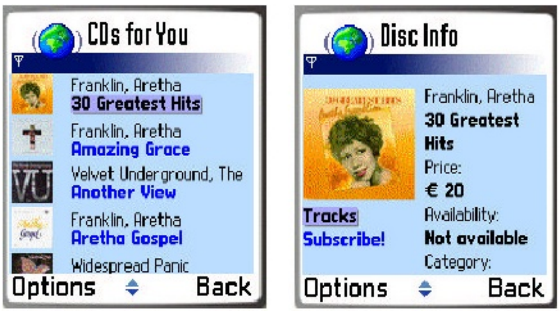


Fig. 2. Disc list and detailed information

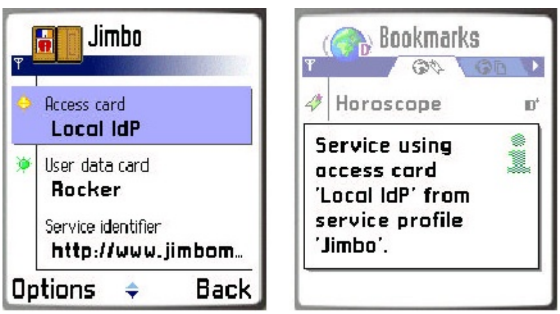


Fig. 3. Service profile and authentication

an authentication request. The screenshot on the right side captures the moment when the access card is taken into use; the notification is received, because the service profile has been configured so (yellow diamond beside the card name in the left side screenshot); possible choices are: *silent*, *notify* and *confirm*. Services are identified by means of their *provider ID* in service profiles (Liberty).

Attribute exchange is implemented based on the principles of the Liberty attribute exchange scheme. Attributes provided by the terminal include nickname,

Fig. 4. User profile: terminal and shop

date of birth, gender, music preferences, country, language, and are defined in *user data cards*.

The screenshot on the left side of Fig. 4 shows how attributes are defined in a user data card. The other screenshot shows what the service can see from the attributes. Note that not all attributes are necessarily sent to the services, e.g. the date of birth is not sent. Each attribute has a *privacy level*, and each service profile is associated with a *sharing level*. The possible privacy levels are: *basic*, *medium*, *detailed* and *private*; sharing levels are the same, except that private is not possible. Only attributes with a privacy level not exceeding the current sharing level are disclosed (private being the highest level).

Payment is done by issuing a signed electronic check. The server at the shop generates the check and then sends it to the client for acceptance and signing. After the purchase, the shop sends the check to the teleoperator who then debits the user's phone bill with the amount.

Privacy is protected by a series of instruments. Connection establishment is controlled by the status attribute of service profiles, which can be: *approved* (connection from service allowed seamlessly, i.e. the browser is started without asking the user), *rejected* (connection from service blocked), *enabled* (confirmation is asked from user) and *disabled* (connection blocked for that day, but enabled again from the following day). The authentication is *pseudonymous* following from how Liberty ID–FF works, i.e. the shop identifies the user with a locally unique identifier. Alternatively, a service profile may define *anonymous* connection by linking to no access card. Disclosure of personal information is controlled by means of the *privacy levels* in the user data cards and the *sharing level* in service profiles, as described earlier (Fig. 4).

3.2 MIDlet over Bluetooth

Content is delivered by means of a MIDlet. The MIDlet itself is also delivered over Bluetooth. Fig. 5 shows the same disc list and detailed disc information pages as their XHTML counterparts (Fig. 2). The similarity is striking. *MIDlet auto–launch*, together with automatic installation, is provided by a middleware developed in–house.

Fig. 5. MIDlet disc list and detailed information

Authentication, *attribute exchange* and *payment* is implemented by means of the same components as in the XHTML version. The MIDlet relays the Liberty and payment requests to the native components and passes the responses received back to the server.

Privacy is protected in the same way as in the XHTML version, given that the same components are used for the privacy–critical authentication, attribute exchange and payment purposes.

4 Conclusions

Building blocks of personalized local services are becoming more and more standard. They "just" have to be put together, but in a way that provides a seamless user experience while maintaining trust. As for the two alternatives, the MIDlet solution proved to be faster and more responsive than the native browser based one and is also less platform-dependent. Overall it provided a better user experience, although it took longer to develop. In addition to content delivery, client auto–launch, authentication, attribute exchange, payment and privacy are also key issues when implementing such services.

Acknowledgements. Special thanks go to Jaakko Lipasti, Nokia Research Center, who has been with us in the project all along and to Seamus Moloney, Nokia Research Center, for the terminal IdP and payment solutions.

References

1. Microsoft .NET Passport. `http://www.passport.com`
2. Liberty Alliance Project. `http://www.projectliberty.org`

Mobility Mapping – A Discount Technique for Exploring User Needs for Future Mobile Products and Services

Val Mitchell[1], Susan Harker[2], and Ken Eason[2]

[1] ESRI: Ergonomics and Safety Research Institute,
Holywell Building, Holywell Way, Loughborough Leics., LE11 3UZ UK
v.a.mitchell@lboro.ac.uk

[2] Dept. of Human Sciences, Loughborough University, Loughborough, Leics., LE11 3TU
{s.d.harker,k.d.eason}@lboro.ac.uk

Abstract. This paper describes a 'low cost' or discount technique termed 'Mobility Mapping' designed to explore user needs for future mobile products and services during concept design and user requirements specification. It can be used to help designers and human factors practitioners encapsulate in scenario form the variability in usage patterns and usage contexts that must be considered within mobile product and service design. Mobility Mapping can also be used to complement field studies by providing a framework for stretching consideration of the context of mobile product use beyond the normal temporal and practical bounds of observation and shadowing activities. This paper introduces the technique and the rationale behind its development.

1 Introduction

"A major problem in exploring user requirements for mobile communication and personal organisation devices is the versatility of usage patterns and usage contexts in which usage takes place." [1] Capturing this versatility or diversity is essential to the design of usable mobile products and services. 'Mobility Mapping' uses schematic maps to provide a visual representation of the 'versatility' of each participant's mobility and social communication context that is understood and shared by both participant and researcher. It is a low cost technique developed to extend exploration of user needs for future mobile products and services during the early stages of design beyond the temporal and physical limitations associated with directly observing mobile users. It helps those involved in design understand the users' world by providing a means of indirectly participating in that world. It also aims to help users relate product concepts to their own experience and therefore bridge the gap between current experience and future use. Mobility Mapping seeks to provide a framework for indirectly exploring the participants' current and potential future use of mobile products in relation to their mobility, their social communication needs and their current use of other non-mobile communication and information technologies. Two freehand maps are produced during Mobility Mapping, a Mobility Map and a Social Communica-

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 476–480, 2004.

tions Map. Figure 1 presents an example Mobility Map. The format of the Social Communications Map is very similar.

1.1 Theoretical Background to the Method

The development of Mobility Mapping was influenced by previous work by Iacucci et al [2]; the use of graphical representations of work context within systems design [3] and Mobile Informatics research [4]. The maps are designed to represent the variability of social and physical contexts within which participants currently use mobile products and therefore encapsulate the variability in needs resulting from the diversity of use contexts. Underpinning the structure of the Mobility Map are the modes of mobility identified by Kristoffersen and Ljungberg [4] adapted to suit representation of mobile lifestyles rather than mobile work. Although providing a limited representation of mobility, the modes are suitable for creating the map structure and therefore prompting wider consideration of use context. Mobility Mapping uses the map drawing process to stimulate study participants to relate how they communicate with others in relation to specific journeys and events they have experienced and how their communication needs are influenced by their mobility and the nature of their relationships.

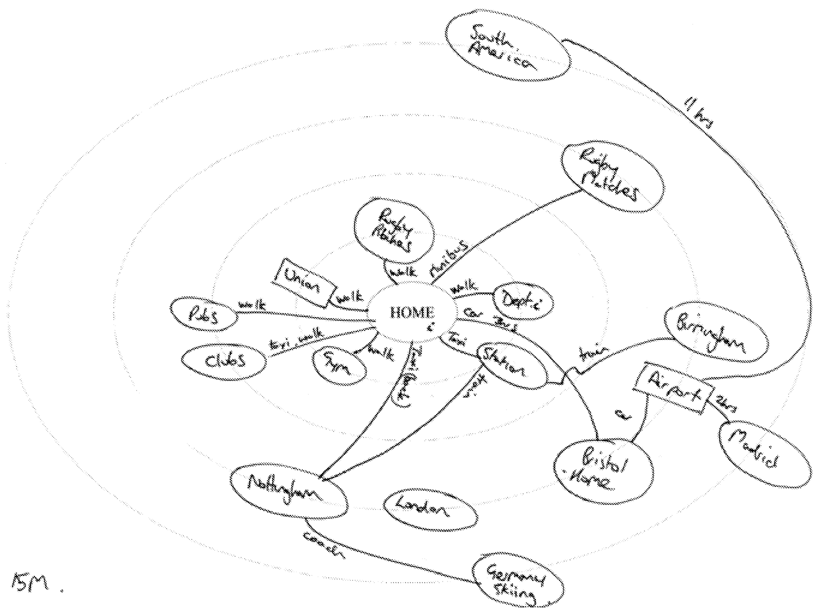


Fig. 1. An example Mobility Map

Mobility Mapping unlike many other techniques for probing the culture and context of use is deliberately systematic, prompting participants to respond within predetermined parameters reflected in the structure of the maps. These relate to factors that

have been identified as playing an influential role in shaping the use of mobile communications - the mobility of the participant, the nature of their social relationships and the availability of other communication technologies. Wood [5] reports, "that individuals select a limited number of all possible associations for report in any given interviewing context." By providing a structure for questioning centered about the personal context of the individual, the maps are designed to stimulate wide recall of use contexts and social relationships.

Within CSCW and related design approaches it is common to use schematic and freehand representations of work in order to build a representation of work context that is understandable to both users and researchers (e.g. Rich Pictures [3] and Prompted Reflections [6]). Mobility Mapping extends this concept beyond workplace systems to provide a means of representing aspects of the context of use relevant to the design of mass-market mobile products. In line with the Rich Picture approach and in contrast to Prompted Reflections, which adopts an ethnographic inspired approach to data collection, the map notation is predetermined although the terminology and level of detail provided is driven by the participant's responses. Dearden and Wright [7] found that the use of rich pictures when studying work context broadened consideration of both the organizational and historical contexts within which work took place. Mobility Mapping similarly allows consideration of a much broader range of events; journeys and places, spread across a much longer period of time than would be feasible within observational based field studies.

Central to the development of Mobility Mapping is the need to find a way for participants to be stimulated to consider their needs for future products in relation to their own personal context. This requires an approach that encourages participants to form a recollection of their own context that is as vivid as possible. Iacucci et al [2] used a map of locations within a game-playing environment to prompt exploration of user needs in relation to context. The map-based environment allowed the participants to consider a range of use contexts simultaneously. Mobility Mapping similarly seeks to create a holistic representation of use context. However the context in this case is personal to each participant whilst in Iaccuci et al's role playing game the context was developed by designers from an assimilation of data collected within a previous ethnographic study. Iacucci et al's SPES technique reported in the same paper enables product concepts to be jointly explored by the participant and researcher within the participant's actual context of use. However shadowing mobile users is fraught with difficulties particularly if the user is extensively mobile or engages in activities or conversations that they do not want observed [8]. Mobility Mapping is designed to complement data collected in situ by prompting participants to consider a wider range of contexts covering over a longer period of time.

2 Mobility Mapping

One or two schematic maps are produced within a Mobility Mapping session. These are drawn freehand on A3 paper pre-printed with a grid of concentric oval rings. The data for the maps is elicited from users through using semi structured interview ques-

tions and card sorting exercises. The ‘Mobility Map’ represents journeys and events experienced by the user. The ‘Social Communications Map’ represents the social relationships of the individual and the media currently used to support these relationships. Both maps are not always needed. For example when exploring user requirements for m-commerce services the Mobility Map was sufficient. Information for the Mobility Map is generated by prompting the participant to recollect firstly places they had visited that day then during the last week and finally to recollect special journeys made over the last year e.g. holidays and trips to visit friends and family.

The *process* of constructing the Mobility Map serves to trigger the user’s recollection of journeys made and places visited in order to provide a wide range of use contexts for considering existing and future mobile product use. It also helps the practitioner understand the mobility patterns of the users and the culture of product use through encouraging the users to recount their past experiences. Constructing the Social Communication map similarly serves to prompt recollection of current use of available communications media and social contacts beyond the most recent and typical. How the user employs the communication media available to them to meet their communication needs is discussed in relation to the different social contexts represented by the maps. The map construction process and subsequent discussions serve to build a shared understanding of the user’s current social network and mobility. When used after direct observation of user behaviour, Mobility Mapping can serve to anchor observational data within a wider framework. Alternatively if used at the beginning of the design process, the technique is used identify unmet communication needs that can then be used as the focus for solution generation. In both cases the aim is to preserve the versatility of the mobile context of use when generating and representing user requirements.

The combination of a narrative generated by the user in relation to a specific context of use naturally leads to the generation of scenarios. These scenarios can be used to ensure the variability of the culture and context of use is communicated throughout the product design process. Mobility Mapping can also be used to facilitate evaluation of emerging product and service concepts. Technology-based scenarios and low fidelity prototypes have been used in conjunction with Mobility Mapping to help users evaluate future products and services concepts in relation to their own personal context of use. The participants use the maps to develop personal use scenarios depicting how they envisage using the concept designs to support their existing social relationships or to meet information or transaction needs experienced during past journeys or events.

3 Conclusion

Mobility Mapping was developed within a research context in order to explore user requirements for future 3G and 2G+ mobile services. It has subsequently been used to elicit user requirements for m-commerce and location based services [9].

Nardi [10] argues that the problem when generating scenarios to inform design is “to find a set of techniques that produces good data reflective of user experiences, yet

practical enough to deploy in everyday settings." Creating data that is reflective of user experiences when designing mobile products and services is particularly problematic because of the diversity of use contexts created by being mobile. The Mobility Mapping technique is low cost, paper based and quick and easy to use. It offers a practical way to encapsulate this diversity either as a means of extending the scope of field based studies or as part of a Participative Design process.

Acknowledgement. The authors are indebted to TTPCom Ltd. for funding this research.

References

1. Väänänen-Vainio-Mattila, K., Ruuska, S: Designing Mobile Phones and Communicators for Consumer Needs at Nokia. In: Bergman, E.: Information Appliances and Beyond. San Diego, Academic Press (2000)
2. Iacucci,G., Kuutti, K. et al.: On the Move with a Magic Thing: Role Playing in Concept Design of Mobile Services and Devices. Designing Interactive Systems: processes, practices, methods and techniques, Brooklyn, New York, ACM Press (2000)
3. Monk, A., Howard, S.: The Rich Picture: A tool for reasoning about work context. Interactions March & April (1998) 21 - 30
4. Kristoffersen,S., Ljungberg, F.: Representing Modalities in Mobile Computing: A model of IT use in Mobile Settings. Interactive Applications of Mobile Computing (IMC'98), Rostock, Germany (1998)
5. Wood, L.E.: The Ethnographic Interview in User-centered Work/Task analysis. In: Wixon, D., Ramsey, J: Field Methods Casebook for Software Design
6. Kensing, F.:Prompted Reflections: a Technique for Understanding Complex Work. Interactions January and February (1998) 7 - 15
7. Dearden, A., Wright, P.: Experiences Using Situated and Non-situated Techniques for Studying Work in Context. Human Computer Interaction – INTERACT'97 (1997)
8. Väänänen-Vainio-Mattila, K., Ruuska, S: User Needs for Mobile Communication Devices: Requirement Gathering Analysis through Contextual Inquiry. Proceedings of 1st Workshop on HCI with Mobile Devices, Dept. of Computing Science, University of Glasgow (1998)
9. Bayer, S., Ross, T., May, A.: The Requirements for Location-based Services: Differences between Target User Groups. 5th Wireless World Conference 15- 16 July 2004, University of Surrey (2004)
10. Nardi, B.: Some Reflections on Scenarios. In: Carroll, J.M.: Scenario based Design: Envisioning Work and technology in System Development. Wiley & Sons Inc. (1995)

Memojog – An Interactive Memory Aid Incorporating Mobile Based Technologies

Kenny Morrison[1], Andrea Szymkowiak[2], and Peter Gregor[1]

[1] Department of Applied Computing, University of Dundee, Dundee DD1 4HN, Scotland
{kmorrison, pgregor}@computing.dundee.ac.uk
http://www.computing.dundee.ac.uk/projects/Memojog

[2] Division of Psychology, School of Social and Health Studies, University of Abertay, Bell Street, Dundee DD1 1HG, Scotland
andrea.szymkowiak@abertay.ac.uk

Abstract. Memory problems are often associated with ageing and are among the most common effects of brain injury. Such problems can severely disrupt daily life and put huge strain on family members and carers. Electronic devices have been used successfully to provide short and timely reminders to memory-impaired individuals. The Memojog project has developed and evaluated a mobile, interactive communication and memory-aid system with elderly and memory-impaired people. The system utilizes current and easily available technology such as the internet and GPRS mobile telephony. This paper will look at the design as well as the successes and limitations of the Memojog system.

1 Introduction

Independence is a defining attribute of being an adult. We guard and preserve our right to determine how and where we live our lives and the routines that we maintain. Loss of this independence is not only demoralizing for all the people directly and indirectly involved but also increases care costs to the individual, their families and the wider community [1]. One condition that can affect the degree of independence is the extent to which people can remember to carry out simple everyday tasks. Memory impairment is among the most common effects of brain injury and is significantly associated with the aging process [2].

Electronic memory aids, such as pagers, dictaphones, mobile phones and small handheld computers or Personal Digital Assistants (PDAs) have been successfully used to display timely and active action prompts to people with prospective memory problems, either in text or speech form [3]. Typically with electronic aids, either the user or user's carer enter data directly onto the device, or they can contact a centre where reminders are entered into a central paging or calling system, and then transmitted to the memory-impaired user at appropriate times. Ease of interaction is critical for user acceptance, and the NeuroPage pager [4], for example, simply required the touch of a button to acknowledge a reminder [1]. Despite the success of NeuroPage due to its simple functionality and its subsequent setup as a commercial ser-

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 481–485, 2004.

vice, it has limited functionality and the necessity of using a commercial paging company can be costly. It is also a one-way communication system and it is therefore not possible to monitor whether users have received and acknowledged reminders.

The ability for the memory-impaired user to interact with a memory aid would greatly enhance its value, particularly for those without widespread cognitive impairment. Recent progress in appropriate technologies, e.g. palmtop computers with telecommunications ability and mobile phones with touch-screen interfaces, make this a possibility.

2 Memojog

The Memojog memory aid system, developed at Dundee University, is a remote and interactive communication system that provides a prompting device for memory-impaired people. A PDA with mobile telephony is used to deliver text-based prompts announced by an alarm to users. The provision of short cues is usually sufficient rather than a detailed description of the action to remind users of the task [5].

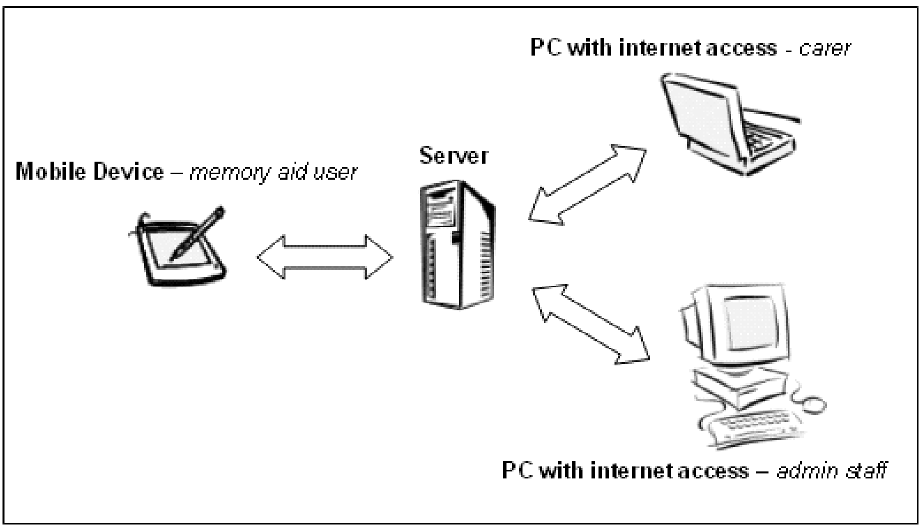


Fig. 1. The architecture of the Memojog system.

In 1998 in the United Kingdom 11,000,000 people were aged over 65 and by 2030 these figures are expected to rise 14,000,000 [6]. Not all of these people will suffer from the memory problems associated with old-age. However, if elderly people could use and become comfortable with an aid such as the Memojog device before any prospective memory-problems occur, they could maintain independence of professional carers, which in turn may result in reduced NHS costs.

Memojog's user prompts or reminders are stored in a remote server and transmitted to the user's device using mobile telephony. The users themselves can add re-

minders directly onto the device or carers/administrators can add reminders remotely using any internet accessible device. Reminders can also be phoned in by users/carers to admin staff or left as messages on an answer machine. This gives users/carers control over the memory aid system, and also provides the option to involve a third party for data entry if they wish to do so. The prompts or reminders are then wirelessly transmitted from the internet database to the PDA at the appropriate times (see Fig. 1 for the system architecture). The system can monitor users' response to reminders and in case of crucial reminders such as 'Take medication' can act accordingly. If a crucial reminder has not been acknowledged, carers/family members can be contacted by text message, prerecorded phone messages or email automatically sent by the central server.

The main function of the Memojog device is to give action prompts to memory-impaired users. However, other functions such as entering reminders or looking up entries for selected days were implemented. In addition, users could look up information on people such as their birthdays or addresses. It was anticipated that users would use the reminder function predominantly, with all other functions learnt with usage of the device. The device can also be used to store other information such as contacts' information, their pictures and details of any appointments with such people. New versions of the software could be developed for different software platforms and devices such as mobile phones with touch screen capability. The digital camera capabilities of some current mobile devices could be incorporated in the design of future systems e.g. the user could take a picture of a person and then add their details and then include it in their list of contacts.

Fig. 2. Two example interfaces

Given the hardware limitations of small handheld devices, such as reduced display size, this can present a major design challenge (Fig. 2). Also, an elderly user group may experience declining visual acuity, contrast sensitivity and reduced sensitivity to

colour, particularly blue-green tones [7]. A specific website aimed at the small PDA screen size was also designed for data entry on the devices.

The PDA that used in the user trials as a memory aid was a Siemens SX45, running the Windows CE 3.0 Pocket PC operating system. This PDA is equipped with a touch screen (a non-reflective TFT LCD with 65,536 colours), the size of which is 240 x 320 pixels (approx. 60 x 78 mm). The dimensions of this PDA are 124 x 87 x 26 mm and it weighs about 300 g. It was decided that four major functions should be clearly visible on the default display: 'View Today' – a list of the current day's reminders, 'View Month' – view reminders from any day, 'Diary Info' – users phone book, appointments etc. and 'Modify Diary' – users can add reminders. A menu structure was not implemented to avoid users becoming disoriented on the small display.

3 Memojog User Evaluation

The system was evaluated with a group of elderly and memory-impaired users at the Oliver Zangwill Centre in Ely. The selection criteria included that participants experienced progressive degenerative disorder problems but no major aphasia, visual problems, or severe general intellectual impairments. There were two field evaluations comprising of 6 participants in each evaluation. Different participants were used in each evaluation and were referred by health care professionals. The participants used the device for 12 weeks, after which the device was withdrawn again. These evaluations have been described in more detail elsewhere [8] and demonstrate that the users were receptive to the system and could use the device easily. Clients commented on the ease of use and the versatility of the device. They particularly like the different functionality the device provided (e.g., diary, phone book, appointments). This illustrates that clients do want additional functionality in a memory aid. Clients also commented positively on the hardware (e.g., "It is easy to see what is on the screen while still being lightweight enough to carry.").

A number of negative comments related to coverage problems e.g. the inability to connect to the relevant website for entering data resulted in frustration to the clients and their carers. While attempting to use something they are unfamiliar with, when problems occur e.g. a non-responsive device, the client/carer is often unsure as to whether they are doing something wrong or not.

Two out of the three major United Kingdom network providers failed to provide reliable, 'always on', GPRS (Global Packet Radio Service) coverage in the area surrounding and including Cambridge, England. Dial-up GSM (Global System for Mobile Communication) coverage was therefore applied which proved to be a more reliable, though more expensive, alternative. In addition, network maintenance procedures conducted by the service provider also posed an impairment of coverage.

The users' comments also expressed some hardware-related problems. The touch screen was not sensitive to button presses and occasionally they had to tap the screen a number of times before they could observe a response. Mobile devices tend to have a small screen that requires fine movements when entering a reminder which can be a

problem for users with a tremor. The text size could prove a problem for users with poor vision. The devices also have a limited battery life and needed to be recharged at regular intervals. As the PDAs that were used in the trials were running an early version of the Windows CE operating system, the Memojog software was lost and had to be reinstalled if the PDA's batteries ran out before being recharged.

4 Conclusions

The Memojog system has many advantages over current electronic memory aids: while also being a mobile and discreet device it is an interactive service and can automatically monitor users' responses to reminders and respond accordingly. As the system uses standard mobile equipment and contracts the users can also use the device as a mobile phone and take advantage of mobile services such as text messaging. There is also clearly enormous potential in this area for experimenting with more recently emerging technologies to support independent living by people with memory impairments, such as picture and video messaging and positioning systems.

The biggest challenge that was encountered throughout the project was the lack of reliable network coverage. The clients themselves were mobile and therefore could easily move from one location that provided network coverage to one where that was not the case.

References

1. Wilson B.A., Scott H., Evans J., Emslie H.: Preliminary report of a NeuroPage service within the health care system. NeuroRehabilitation 18: 3-8
2. Huppert F.A., Johnson T., Nickson J.: High prevalence of prospective memory impairment in the elderly and in early-stage dementia: Findings from a population based study. Applied Cognitive Psychology 2000; 14: S63-S81
3. Inglis E.A., Szymkowiak A., Gregor P., Newell A.F., Hine N., Shah P., Evans J.J., Wilson B.A.: Issues surrounding the user-centred development of a new interactive memory aid (in press in International Journal of Universal Access in the Information Society, special issue "Countering Design Exclusion")
4. Hersh N.A., Treadgold L.G.: NeuroPage: the rehabilitation of memory dysfunction by prosthetic memory and cueing. Neurorehabilitation 4: (1994) 187-197
5. Wilson B.A., Evans J.J., Emslie H., Malinek V.: Evaluation of NeuroPage: a new memory aid. Journal of Neurology, Neurosurgery and Psychiatry: (1997) 63: 113-115
6. HMSO Population Trends. Her Majesty's Stationary Office, London (1998)
7. Stephens, S. D. G.: Rehabilitation and service needs. In M. E. Lutman, & M. P. Haggard (Eds.), Hearing Science and Hearing Disorders London: Academic Press. (1983) 283-324
8. Symkowiak A., Morrison K., Shah P., Gregor P., Evans J., Newell A. and Wilson B.: Memojog – an interactive memory aid with remote communiction, Proceedings of the 2nd Cambridge Workshop on Universal Access and Assistive Technology, University of Cambridge (2004)

JIKUKAN-POEMER: Geographic Information System Using Camera Phone Equipped with GPS, and Its Exhibition on a Street

Yasuto Nakanishi[1], Masashige Motoe[2], and Shohei Matsukawa[3]

[1] Tokyo Univ. of Agri. & Tech.
yasuto@cc.tuat.ac.jp
[2] Miyagi Univ.
motoe@myu.ac.jp
[3] 000studio
sho@000studio.com

Abstract. In this paper, we introduce a geographic information system using camera phone equipped with GPS and its exhibitions. We have proposed a new kind of interface to see lost of pictures which have location information, and in the exhibition, we projected our system onto a shopping street in Japan and held it as a photography exhibition. We studied 700 pictures sent for the exhibition and three peculiar motifs were found.

1 Introduction

In the fields of town management and urban planning, the GIS (geographic information system) aimed at helping citizens to participate in making a city master plan or to exchange information among communities has been developed [1-2]. Users in these systems annotate physical spaces with text notes and photos, and share information in the real spaces, and exchange information. Some location-based systems allow users to participate as content providers for making social and dynamic information spaces [3-5], and users annotate text notes to physical spaces utilizing PDAs, and this allows users to submit information where they want to do it. However, in these researches, utilized devices need extra devices (a GPS card or a WLAN card) and users to utilize applications developed specially, and users can't annotate pictures. This paper introduces our GIS that utilizes camera phone equipped with GPS. Our purpose is same as [1-6], however in our system, users annotate not only texts but also pictures by sending an email from camera phone. When it contains a lot of pictures, it might not be appropriate to show each photo according to its location information accurately because some photos will be overlapped and it will be difficult to see a map below pictures. These problems are true for other GIS or various "moblogging" systems those post emails from camera phone to "blogging" homepages or map them onto a map. To cope with these problems, we have proposed a new kind of interface to see lots of pictures in parallel with fade-in and fade-out. This paper introduces its exhibition on a shopping street and our study of 700 pictures those were sent for it.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 486–490, 2004.

2 System

Our system consists of a mail server, an email client developed with JavaMail API, a WWW server (Tomcat4.1.2) with a database (MySQL3.23.52), a Java Servlet and a viewer client developed with Macromedia Flash (Fig. 1). As a content provider, a person sends an email with a picture and location information attached to a destination email address which was decided beforehand. The location information precision of the GPS in mobile phone is about 10m. Then the email-client receives sent emails every one minute, and then obtains the email address of the user, the subject, the content text, the sent time, the latitude, the longitude and the attached picture from the received email and stores them into the database. The viewer-client sends a query which includes times, longitudes and latitudes to the Servlet every two minutes periodically. Then the servlet sends a query to the database, and returns the searched result to the viewer-client. After receiving the searched result, the viewer-client parses it and begins to download pictures from the WWW server, and shows pictures using location information on a map.

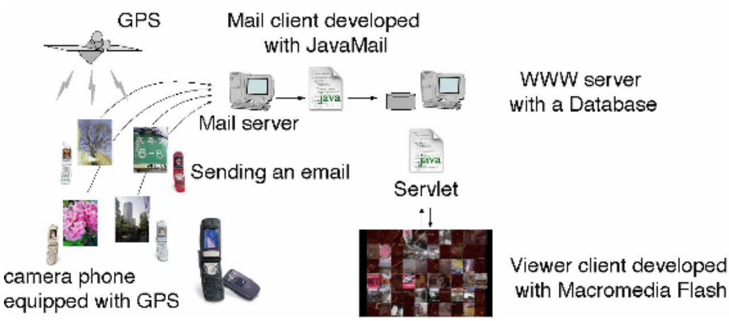


Fig. 1. The system architecture in our system

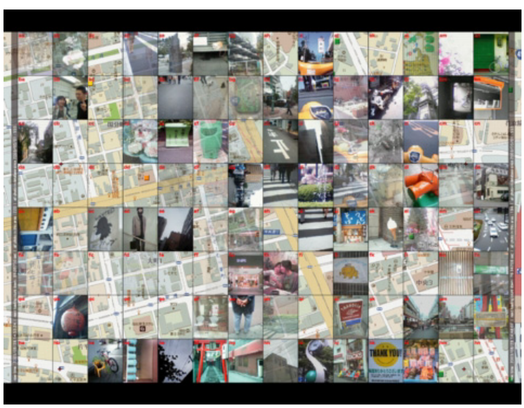


Fig. 2. The viewer-client in our system: showing photos from camera phones on a map according to location information using a grid system.

Our current viewer-client is assumed to show several hundreds pictures, and we have proposed a viewing interface that utilizes a grid system in order to cope with the problems described in the section 1. In our system, each picture is mapped into one cell of the grid according to its location information. When a cell contains some pictures, it manages them with using a list sorted by the sent time. Each cell shows pictures sequentially with letting each picture fade in and fade out, and this process is done in each cell in parallel. This enables us to see lots of pictures without overlapping and to see a map in an interval of a fade-out and a fade-in (Fig. 2).

3 Exhibition on a Street

We introduced our system on a shopping street in Sendai where is a large city in Japanese Tohoku region from 28/05/2003 to 01/06/2003. The client was projected onto the street using two projectors those were set on the shopping arcade (Fig. 3). We put a screen of 5.4m x 3.6m. It is made of retro-reflective sheet, which consist of thousands of precise prism particle per square inch and have superior reflection ability. The members of the photography club in Miyagi Univ. took and sent pictures to our mail server. We selected them as initial content providers because their pictures might provoke contributions from general persons. They usually take pictures with single-lens reflex cameras and have also general exhibitions in galleries with putting their photographs in frames. Some members have mobile phones equipped cameras, however, it was the first experience for them to hold a photography exhibition using cameras in mobile phones and to have an exhibition on a street. They used the mobile phones equipped with cameras of 1.5M pixels and we showed about 700 pictures sent from mobile phones. The pictures were taken within a 1.5km x 1km area. However most of emails were sent beforehand by the photography club members, some visitors sent emails with using their own mobile phones in the exhibition period. This fact would show the merit that the utilized devices in our system are popular consumer products. The street is one of the most active shopping district in Sendai and its location was in front of a long-established department store, and more passers-by watched it than persons who visited in order to watch it. They enjoyed seeing pictures sent from mobile phones while walking on the map. Such experiences, which are different from watching the client with a PC monitor, might encourage social

Fig. 3. Exhibition on a shopping street in Sendai.

communications between users more actively. Estimating from the number of the leaflets that we handed, the number of the persons who heard our explanation is approximately 1000. Old and middle age persons tended to admire the technical features in our system which utilized location information by GPS and updated shown pictures automatically. On the other hand, persons from 10's to 30's tended to have interests in expanding the potential of sharing pictures mutually because sending pictures from camera phone is very popular for them. Some persons asked us if they can send pictures taken with their own mobile phones or if they can send their own favorite pictures stored in their own mobile phones and some of the did it. Such reactions were peculiar to young generations. This difference would be resulted from the fact that sending emails with pictures from mobile phones is a daily activity to most of young persons but is unusual for most of old and middle age persons.

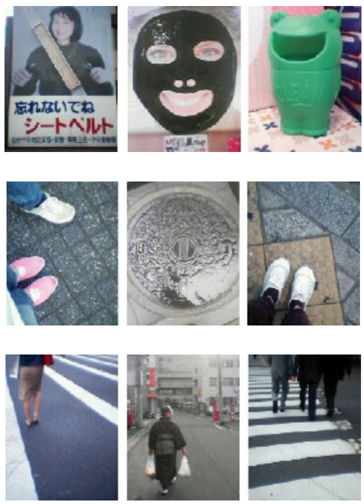


Fig. 4. Three motifs which would be peculiar to our exhibition in Sendai, “funny one” (top), “ground” (middle) and “one’s back” (bottom).

We studied the motifs of the 700 pictures and had interviews with the members of the photography club. We compared the motifs taken by cameras in mobile phones with ones taken by single-lens reflex cameras those they use usually, and we found three motifs peculiar to mobile phone cameras. They are "funny one", "ground" and "one's back" (Fig. 4). "Funny one" means the pictures which took strange signboards or funny objects as they are without photogenic compositions. Such an activity will be caused from the feeling that they want to show such pictures to others naturally and to share such discoveries. In taking a picture and sending it to friends, a camera phone is quite easier than a general digital camera or a single-lens reflex camera, and this fact might had promote taking "funny one". The "ground" means pictures which took own feet, zebra crossings or manholes with turning cameras to the ground. The members of the photography club told that they had never taken pictures of such compositions in single-lens reflex cameras. The restriction of adding location information in taking pictures seemed to have aroused interest in the ground which is the coordinate plane to prescribe the place. The "one's back" means that there are few pictures taking a

person from the front, however, there are many pictures taking a back of a person. As a camera phone becomes popular, it has been worried that cameras in mobile phones might be used for taking a picture furtively. Some people show wariness when a stranger aims a camera in a mobile phone to them, and according to the members of the photography club, more people seem to show such wariness than when a single-lens reflex camera is aiming. Therefore it would be difficult to take a picture of someone else from the front, and as the result, pictures taking someone's backs seemed to increase.

4 Discussion and Conclusions

In this paper, we introduced our GIS using camera phone equipped with GPS and its exhibition on a street with a large horizontal screen. Such a spatial and public information space showing linked to the physical space would bring more chances of conversations between citizens than PC monitors or Plasma Display Panel. When walking on our map, most persons seemed to think that some interactions would occur by stamping the screen or according to the location of person. We don't have developed such interactions, however, we will bring some interactions in the near future. Embedding sensors in the floor screen or putting location sensors such as image-processing system above the screen would be possible methods. A camera in a cellular phone is different from a general digital camera in the point that it is equipped to send pictures to friends or acquaintances. When people send a picture from a camera phone, they would be motivated with emotions that they want to introduce their impression as a short message and a small picture. We studied about 700 pictures sent for our exhibition and found three motifs which would be peculiar to our exhibition. These motifs would bring discussions on a new photography in city spaces and on a new viewpoint to city spaces which mobile devices have brought us. Our system is developed not only for photograph exhibitions but also for supporting communications between citizens. We held workshops about town management in a city of the suburbs of Tokyo and will evaluate how our system can promote information exchanges among citizens and stimulate communications between citizens and bureaucrats of the city.

References

[1] Kakiko Map, http://upmoon.t.u-tokyo.ac.jp/kakikodocs/

[2] Town Resource Database, http://www2.info-mapping.com/sendai/ index.html

[3] Burrell, J., Gay, G.K., Kubo, K., and Farina, N., Context-Aware Computing: A Test Case, Proc. of 4th Int. Conf. on Ubiquitous Computing, pp. 1-15 (2002).

[4] Espinoza, F. et al., GeoNotes: Social and Navigational Aspects of Location-Based Information Systems, Proc. of 3rd Int. Conf. on Ubiquitous Computing, pp. 2-17 (2001).

[5] Tarumi, H., Morishita, K., and Kambayashi, Y., Public Applications of SpaceTag and their Impacts, Digital Cities, Springer LNCS, Vol. 1765, pp. 350-363 (2000).

[6] Ananny, M., Biddick, K. and Strohecker, C., Constructing Discourse with Ethnographic/SMS "Texts", Proc. on Mobile HCI 2003, Springer LNCS Vol. 2795, pp. 368-373 (2003).

Visual Object Detection for Mobile Road Sign Inventory*

Christin Seifert[1], Lucas Paletta[1], Andreas Jeitler[2], Evelyn Hödl[2], Jean-P. Andreu[1], Patrick Luley[1], and Alexander Almer[1]

[1] JOANNEUM RESEARCH Forschungsgesellschaft
Institute of Digital Image Processing
Wastiangasse 6, A-8010 Graz, Austria
lucas.paletta@joanneum.at
[2] FH Joanneum GmbH
University of Applied Sciences
Alte Poststrasse 149, A-8020 Graz, Austria
andreas.jeitler@fh-joanneum.at

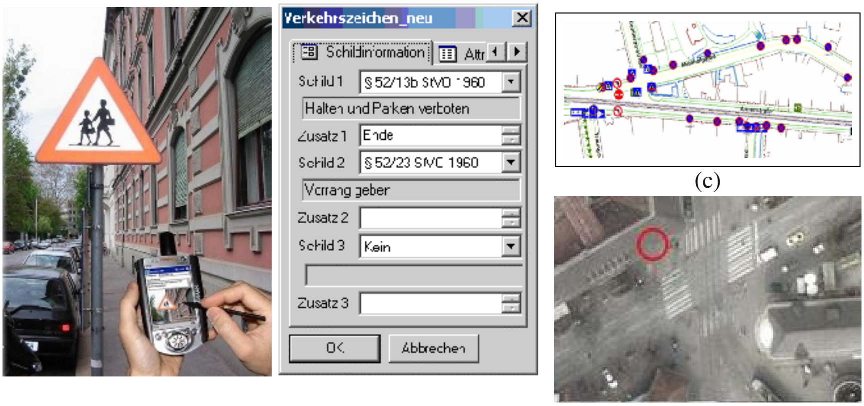
Abstract. For road sign inventory and maintenance, we propose to use a mobile system based on a handheld device, GPS sensor, a camera, and a standard mobile GIS software. Camera images are then analysed via object recognition algorithms which results in an automated detection, i.e., localisation and classification of the signs. We present here the localisation of points and regions of interest, the fitting of geometrical constraints to the extracted set of interest points, and the matching of content information from the visual information within the sign plate. From the preliminary operational state of the vision based road sign detection system we conclude that the selected methodology is efficient enough to achieve the requested high quality in object detection and classification.

1 Introduction

Road sign inventory and maintenance is today performed on the basis of GIS (geographic information systems) relying on geo-reference and content information about the signs. For the registration of previously undocumented sign objects, we propose to use a mobile system based on a handheld device, GPS sensor, a camera, and a standard mobile GIS software. The camera captures the appearance of the signs, images are then analysed via object recognition algorithms which results in an automated localisation and classification of the signs. Automated classification enables faster and more reliable processing than manual interaction with a GUI. We present here the full process chain for a robust recognition of sign objects: the localisation of regions of interest (ROI), the fitting of geometrical constraints to the ROIs, and the analysis of visual information within the sign plate. From the preliminary operational state of the mobile vision based inventory system we conclude that the selected methodology is efficient enough to achieve the requested high quality in object recognition.

* This work is funded by the European Commission's IST project DETECT under grant number IST-2001-32157, and the Austrian Joint Research Project 'Cognitive Vision', sub-projects S9103-N04 and S9104-N04.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 491–495, 2004.



(a) (b) (c) (d)

Fig. 1. (a) Mobile inventory system, (b) manual sign definition interface, (c) sign symbols within ArcPad plan, (d) air-borne images for positioning.

2 Mobile Inventory System

The goal of the system is to attain efficient acquisition, attribution, and localisation of road signs for the purpose of inventory, inspection and maintenance. The system is based on ArcPad (CE Software) on a handheld device, GPS localisation, and a vision based module for object recognition. The mobile prototype was developed on a Pocket PC (HP's iPAQ 3970) in connection with a blue tooth GPS receiver (Fortuna BTGPS). A digital image of the road sign is captured by the mobile user within the urban/rural environment. The PDA based ArcPad application supports the data acquisition (Fig. 1a) and overtakes the central data storage. It allows automatic gathering of parameters like GPS position, current time stamp (UTC) and also the user ID. The acquisition of field data is assisted using a "moving map" functionality based on GPS location and accordingly transformed handheld position into a current map projection. The uncertainty in the positioning is compensated by a 'snap-to-point' method, in case the location of the road sign would be known beforehand. We will use standard auxiliary tools for geo-referencing, e.g., indication of location on maps or airborne image data, etc., Fig. 1d. It is planned to integrate the camera into the mobile device, and to outline a client based object detection module. System enhancements of 'mobile vision' capabilities will allow to automatically gather road sign attributes, which can be supplemented and also changed by the user, and will also perform validity and plausibility tests on the acquired data. For the detailed attribution of the collected data, we currently apply an off-site post-processing, consisting of, (i) geo-referencing of road signs via time synchronisation between digital image and track log data of the GPS receiver, and (ii) object detection of the automated identification of every road sign in the image.

(a) o_1 (b) o_2 (c) o_3 (d) Content

Fig. 2. Road sign subwindow detections (a)-(c) and associated sign plate content (d) corresponding to objects in (a)-(c).

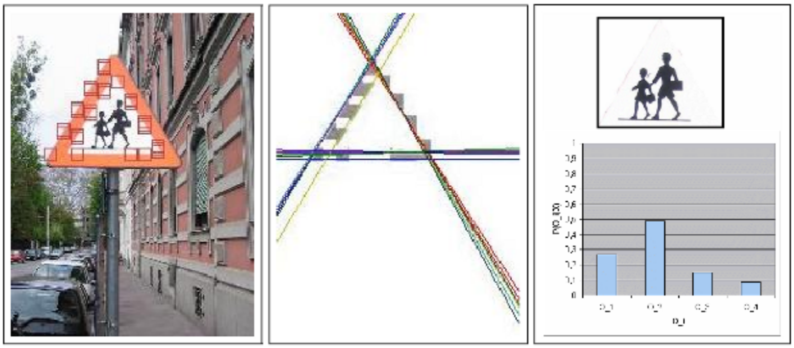
3 Visual Object Detection

The goal of the object detection module is to localise and identify road signs within the captured images. The images can be taken from a camera attached to the handheld, or a connected digital camera. The image should then be either transmitted to a server, or directly processed on the mobile client. While this work considers the first case, the latter is still topic of ongoing research.

The demand on the quality of service is high, in order to represent a beneficial technology for mobile road sign inventory and maintencance, and to substitute any manual interaction. Therefore, we aim at 100 % detection rate without positive falses. Efficient visual recognition of road signs has already been proposed in the context of automotive applications [Gavrila and Philomin, 1999], and been refined with respect to ambiguous information when recognising signs from a moving vehicle, e.g., in [Escalera et al., 2003]. However, all these systems have in common, that (i) they were operated on images taken from a moving vehicle, under bad weather conditions, or just from a far distance, and accordingly, (ii) the recognition rate was 'clearly below' 100 %. In the proposed application, (i) the mobile user has an impact on image acquisition (using flash and appropriate distance to object, (ii) we are required to achieve a detection rate above 98% to compete efficiency in manual interaction.

We follow here a framework of cascaded processing on the visual information, in order to become more robust in the interpretation, as follows,

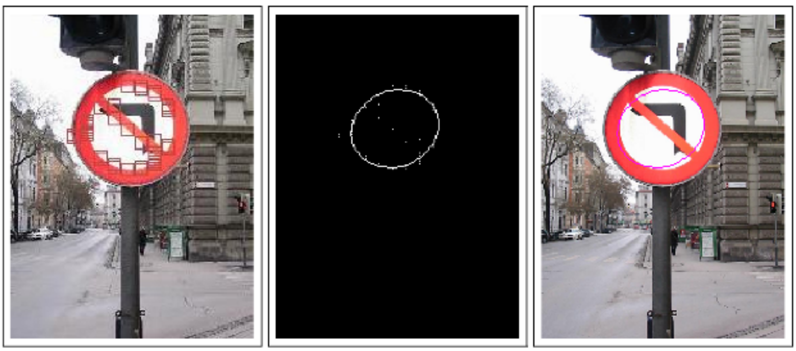
1. **Pixel classification from learned color filters.** We learn color filters from images to classify pixels, applying an EM (expectation maximisation [Dempster et al., 1977]) cluster algorithm and a maximum likelihood classification approach thereafter.
2. **Local regions of interest.** Local subwindows (e.g., 10×10 pixels) are then interpreted for further processing (Fig, 3a,d). We exploit the information from bimodal distributions of color pixels (in analogy to [Matas et al., 2000]), characterising typical color contrasts found in road signs, such as, 'RW' (red-white), 'BW' (blue-white),



(a) Subwindows

(b) Hough transform

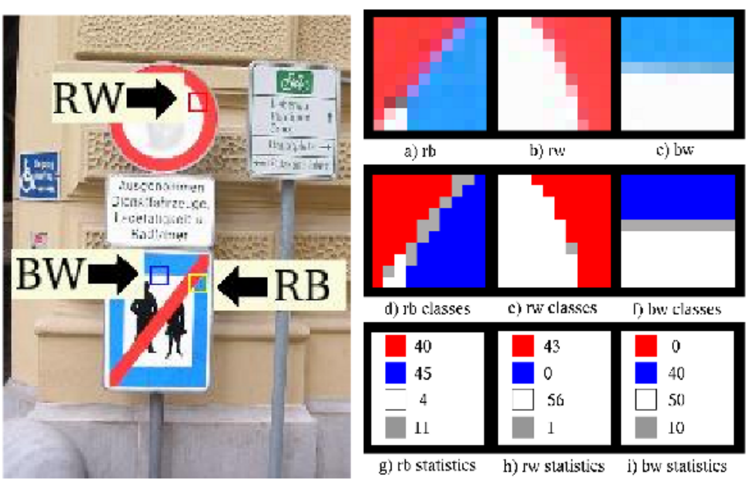
(c) MAP Object hypothesis (o_2)



(d) Subwindows

(e) Ellipse fitted

(f) Ellipse overlaid



(g) Selected subwindows

(h) Subwindow content: a,c) original colors, d-f) color classification, g-i) class statistics

Fig. 3. Cascaded road sign detection.

and 'RB'. We examine a plausability test by applying rules for window classification, such as, 'if W $>$ 35% AND R $>$ 35% then RW'.

3. **Extraction of sign geometry.** We then extract triangle and circular structure from local subwindows. We apply two primitive extractors, (i) an ellipse fitter [Fitzgibbon and Fisher, 1999] and (ii) Hough Transform [Illingworth and Kittler, 1988] to extract straight lines (Fig. 3b,e). The primitives are determined either on center points of the local subwindows. We perform some post-processing, such as, clustering similar lines, rejecting 'useless' lines and ellipses, i.e., those who have not sufficient support from the detected subwindows.
4. **Matching the sign content with prototypical patterns.** The final step is to extract the sign content, and apply a matching (using correlation) to stored prototypical road sign patterns, for road sign identification (Fig. 3c).

The experiments demonstrate that the preliminary system already achieves very robust interpretation of the road sign images. We intend to (i) extend the object database to classify up to 120 Austrian road signs, and to (ii) make the approach more robust by applying a probabilistic framework all over the recognition process. 'Mobile Vision' is an emergent technology in mobile computing, with potential applications in automated translation [Yang et al., 2001] and object based tourist information systems [Fritz et al., 2004].

References

Dempster et al., 1977. Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, 39(1):1–38.

Escalera et al., 2003. Escalera, A., Armingol, J., and Mata, M. (2003). Traffic sign recognition and analysis for intelligent vehicles. *Image and Vision Computing*, 11(3):247–258.

Fitzgibbon and Fisher, 1999. Fitzgibbon, A. W.and Pilu, M. and Fisher, R. B. (1999). Direct least-squares fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):476–480.

Fritz et al., 2004. Fritz, G., Seifert, C., Paletta, L., Luley, P., and Almer, A. (2004). Mobile vision for tourist information systems in urban environments. In *International Conference on Mobile Learning, MLEARN 2004*.

Gavrila and Philomin, 1999. Gavrila, D. and Philomin, V. (1999). Real-time object detection for smart vehicles. In *Proc. International Conference on Computer Vision*, pages 87–93.

Illingworth and Kittler, 1988. Illingworth, J. and Kittler, J. (1988). A survey of the Hough transform. *Computer Vision, Graphics, and Image Processing*, 44:87–116.

Matas et al., 2000. Matas, J., Koubaroulis, D., and Kittler, J. (2000). Colour image retrieval and object recognition using the multimodal neighbourhood signature. In *Proc. European Conference on Computer Vision*.

Yang et al., 2001. Yang, J., Gao, J., Zhang, Y., Chen, X., and Waibel, A. (2001). An automatic sign recognition and translation system. In *Workshop on Perceptive User Interfaces*. ACM Digital Library.

Integrated Care Plan and Documentation on Handheld Devices in Mobile Home Care

Isabella Scandurra, Maria Hägglund, and Sabine Koch

Department of Medical Sciences, Medical Informatics and Engineering
{isabella.scandurra, maria.hagglund, sabine.koch}@medsci.uu.se
http://www.medsci.uu.se

Abstract. Mobile work situations within home care of the elderly require immediate and ubiquitous access to patient-oriented data. We have developed a PDA based prototype that provides both access to the current care plan and an intuitive way for home help personnel to document the performed measures during mobile work. System development was conducted according to a user centered design approach in interdisciplinary working groups consisting of home help personnel, nurses, physicians, medical informaticians, system developers and usability experts. In this paper, we describe how the development of the prototype was performed and present our design considerations as well as the resulting prototype.

1 Introduction

Ageing of population and shortage of resources trigger a trend for decentralisation of in-hospital care towards primary and home health care in the western world. This is especially true for Sweden which has one of the most rapidly ageing populations in Europe with 18% of the population aged 65 or over [1]. The organizational shift of responsibility for the domestic care of the elderly and disabled people from the county councils to the local authorities [2] increases the amount of actual home care performed by the home help personnel (HHP). To guarantee quality-oriented health care for the elderly, the HHP need supporting systems to cope with their mobile work situation. The HHP mainly consist of assistant nurses and they perform certain medical tasks on delegation. The increased responsibilities lead to new requirements regarding care planning and documentation of performed measures. Making a care plan for a patient involves identifying health problems, finding objectives and planning measures to be performed in order to reach the objectives. Ideally the care planning should be performed by the entire team involved in the patient's health care e.g. the district nurse (DN), the HHP, the patient and his relatives. Equally important as the care planning is the documentation of the performed measures, which gives feedback into the care planning process. As documentation is performed in a mobile work situation, it seems feasible to use mobile documentation tools. Hitherto, most mobile IT-based systems for home health care focus on vital sign measurement using differ

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 496–500, 2004.

ent sensor technologies or administrative support for time measurement. Support tools for care planning and mobile documentation only exist sparsely today.

We have been identified mainly two work tasks:

1. documentation of performed measures which consists of a number of mobile activities and
2. care planning which consists of a number of non-mobile activities.

Therefore, we developed two different systems, on different devices with different content and interfaces to support these two work tasks, a PDA-application for documentation of performed measures and a web solution for care planning.

In this paper, we will focus on the PDA application, the design rules for GUI development and navigation aspects. The care planning tool is important to the context of the PDA tool since they are integrated, giving the HHP access to e.g. the planned measures from the current the care plan in their mobile work situations.

2 Method

2.1 User Centred Systems Development

The HHP have rarely been in focus for user-centred systems development. The user needs have subsequently not been sufficiently described and analysed, making a thorough user needs analysis necessary before developing a new IT-support in this area [3]. We perform the user needs analysis, taking the entire work process into consideration. In order to specify the work situations, we use the UCSD [4] approach where interdisciplinary working groups improve and analyse both the work process and the IT tools being developed. The working groups consist of 5 home help workers, 3 district nurses, 2 general practitioners (GP), 3 usability experts and 3-5 system developers.

The design process includes activities such as analysis of the current work, finding essential goals and barriers that obstruct work in order to envision the future work and practices.

For the design of the graphical user interface, we applied the workspace metaphor [5] that describes a way of structuring according to work situations. Instead of working with several applications, the user has a number of workspaces that are carefully designed and customized screens to support the user in the performance of different work situations. The work spaces become interface objects on a top level containing all the information objects needed in a specific work task [6]. We adapted the workspace metaphor, originally designed for desktop working environments to the mobile work situation.

2.2 Technology Used

The Integrated Care Plan and Documentation System is partly implemented as a Pocket-PC application developed in C#, MS Visual Studio .NET environment on HP

IPAQ 5550 and partly as a web application on a desk top PC. The documentation of the performed measures during mobile work is conducted on the PDA and the web application is used for the care planning procedure in the office and at a patient's home.

3 Results

A general problem in the design of usable interfaces concerns how a large and complex information structure can be visualised and controlled efficiently on a relatively small screen. A common solution to the limited screen space problem is to divide an application into a number of different windows, often hierarchically structured [5]. According to Woods & Watts a hierarchical application structure causes navigational difficulties, the user easily gets lost in information space [7]. Our PDA application is not using a navigator page, nor deep hierarchical tree structures but tab views showing, and following, the work processes executed by the personnel on the field.

The workspace metaphor [5] is implemented as tabs where each tab corresponds to a work situation on the display, thereby adapting it to a small screen. Practising the workspace metaphor by using tabs in the application, both a work flow and a work situation-oriented system perspective are achieved.

Three logical levels appear consistently under each tab containing: (1) an *overview*, such as a list of all the items in the object, (2) a *detailed view* showing all the details of a selected item and (3) a *writing mode* where the personnel can still see the information needed when inserting new data (see Fig 1). Many IT systems in health care hamper the work flow today by not providing both read- and write modes in the same view.

Fig. 1. Two levels from the care plan documentations tool. The detailed vie to the left, showing all the details of a selected objective and a writing mode where the personnel can still see the information needed when inserting new data.

Scrolling is allowed since more levels would complicate the conceptual image of the application's interaction flow, and is facilitated by a hardware navigation button.

The tab overview enables the users understand how to interact with the prototype and shows what has to be done in the work process. Relevant information is always on top and the high use activities are just one, or a few, click away

The users state that it is more important to read patient information when in a patient's home compared to the information to report. But when there is something to report an easy input is mandatory [8]. Textual data entries are only required when absolutely necessary, most of the time the personnel can use a quicker way of inserting data into the system, by simply clicking check boxes, option buttons or using drop-down menus.

4 Future Work

Home care of the elderly today is performed by different types of care providers. The main obstacle for quality oriented home care is the insufficient information and communication flow between these care providers. Today, information is documented in several systems, such as the HHP care planning system, the nursing record and the medical record. To really support the team work process, an integrated solution, such as a shared care plan, enabling a reduction of redundant documentation is necessary. We will further work for a seamless and consistent information flow by integrating the prototypes presented in this paper into a virtual health record that will enable each user to access the information needed [9]. Further enhancements of the usability, both for the stationary and the handheld devices, will be provided to the users, e.g. other input facilities such as portable keyboards or electronic paper.

The limited screen space is one of the major obstacles for creating a usable tool based on work situations. This work shows that the workspace metaphor enhances efficiency also on small screens, and the transition of the metaphor to the small screen devices will be further investigated and further prototypes tested.

Acknowledgements. This work has been performed within the project "Technical Support for Mobile CloseCare", sponsored by VINNOVA - Swedish Agency for Innovation Systems (P23037-1 A) and Trygghetsfonden and by its clinical and industrial partners**.**
Malin Gustafsson, student at the department of Information Technology, Uppsala University, did her master thesis within the project and has greatly contributed to the development of the prototype.

References

[1] Swedish Institute. The Swedish Population. FS 67 ISSN 1101-6124, june 2003; http://www.sweden.se/templates/FactSheet____6048.asp

[2] Swedish Institute. The Health Care System in Sweden. FS 76 ISSN 1101-6124, sept 2003; http://www.sweden.se/templates/FactSheet____6856.asp

[3] Scandurra, I., Hägglund, M., Johansson, N., Sandblad, B.,. Koch, S.; User needs for development of context dependent devices in mobile home care , L.Chittaro (Ed.) Mobile HCI 2003, LNCS 2795, pp 446-450, 2003, Springer Verlag Berlin Heidelberg 2003.
[4] ISO 13407:1999: Human-Centred Design Processes for Interactive Systems. International Organization for Standardization, Geneva. 1999
[5] Lif, M., Olsson, E., Gulliksen, J., Sandblad, B. Workspaces Enhance Efficiency – Theories, Concepts and a Case Study. Informations Technology and People, 30(4), 2000
[6] Gulliksen, J., Göransson, B. and Lif, M. A User-Centred Approach to Object-Oriented User Interface Design. In Van Harmelen, M. (ed). *Object Modeling and User Interface Design: Designing Interactive Systems,* Addison-Wesley, US, 2001, pp. 283-312.
[7] Woods, D.D., & Watts, J.C. (1997) How Not to Have to Navigate Through Too Many Displays. In M.Helander, T.K. Landauer, & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (pp. 617-650). Amsterdam: Elsevier Science B.V.
[8] Scandurra I, Hägglund M, Koch S: Visualization and Interaction Techniques for Mobile CloseCare, Proceedings of World Congress on Medical Physics and Biomedical Engineering, Sydney, 2003 ISSN 1727-1983/ISBN 1 877 040 142
[9] Koch, S., Hägglund, M., Scandurra, I., Engström, M., Moström, D.: Towards a virtual health record for mobile home care of elderly citizens, accepted at MedInfo 2004 in San Francisco, September 2004

CENTAUR: A Two-Panel User Interface for Mobile Document Access

Greg Schohn and Adam Berger

Nokia Inc,
503 Martindale Street
Pittsburgh PA 15212
{greg.schohn, adam.berger}@nokia.com

Abstract. This paper introduces a novel user interface designed to mitigate some of the usability problems in mobile web access. The interface consists of two side-by-side panels for representing a richly-formatted document (e.g. a web page). In one panel is a "wide angle" view of the entire document, partitioned into regions. In the other panel is a "zoomed in" view of the currently-active region. We describe a prototype system, called Centaur, which automatically and in real-time reformats electronic documents into this side-by-side view.

1 Introduction

Accessing richly-formatted content on mobile browsers—PDAs and handheld phones, for instance—can be a trying experience. The causes are multifold: limited memory and processing power of the devices, high latency of cellular data networks, and input devices (such as the Bell keypad). Most fundamental, however, is that web content is almost always composed for eventual viewing on a PC display. The difference in display capabilities between a PC and a mobile device is typically significant.

This paper introduces CENTAUR, a prototype system whose purpose is to bridge this gap through the use of a novel user interface. Designed for the emerging class of higher-resolution mobile devices, CENTAUR renders documents using a technique that is a hybrid between two popular approaches:

1. High-fidelity reproduction of the original page format
2. Content reformatted to accommodate the target device's geometric constraints

The recent trend in mobile devices is towards larger and higher-resolution displays. One recently introduced device, the Nokia 7700, is on par with the VGA standard of 15 years ago: 640x320 pixels and 65,536 colors. We can expect this trend to continue, but only up to a point: form factor constraints demand that mobile devices and their displays remain small. Assuming the resolving power of the human eye doesn't improve, there will always exist a biologically-imposed limit on the amount of information presentable on a fixed-size mobile device display. In other words, the gap between PC and mobile display capabilities won't disappear anytime soon.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 501–506, 2004.

Fig. 1. Conceptual views of two commonly used small-screen viewing modes. **Left**: Narrow-screen layout of a web page. The original document has been reformatted into a linear or "ticker-tape" form, which requires only up-down scrolling to view the content. **Right**: Full-scale view. This mode maintains complete fidelity to the original layout, but requires both up-down scrolling and left-right panning

1.1 "Narrow-Screen" and "Full-Scale" Layout

Fig. 1 depicts the two display techniques in common use in mobile browsers.

Systems implementing *narrow-screen layout* (NSL) modify the layout of the original content to fit the width of the device display: adapting or removing tables, frames, multicolumn text, shrinking images to fit the display, and so on. The end result is a document which may be viewed in its entirety using vertical scrolling alone; no horizontal panning is required. There exist several commercial systems implementing NSL. For instance, the Opera Smartphone mobile browser can be configured to render HTML in NSL [7].

NSL's obvious benefit is the ease of navigation through the reformatted page. On the other hand, NSL suffers from what one might call *ill-defined serialization*. That is, the optimal or most intuitive serialization of a two-dimensional web page is not always obvious[1]. In some cases, a meaningful projection onto one dimension does not even exist. For example, should cells from a two-dimensional table be presented in row-major order or column-major order? How should the serialization procedure handle elements which have been assigned an absolute (x,y) position on the page using CSS?

The right half of Fig. 1 displays what we will call *full-scale layout*. Here the content isn't reformatted, but instead rendered as-is in the mobile browser. The user must

[1] In fact the problem is even more severe. Using CSS and dynamic HTML, web authors can provide a third dimension to their web content: pop-up windows, hovering text, layers, and so on.

Fig. 2. CENTAUR view of a web page. When a user selects a region from the left panel, an NSL view of that region appears in the right panel. We call this operation region selection. When a user clicks on a hyperlink in the right panel—thus loading a URL—a thumbnail image of the new document appears in the left panel and an NSL view of the default region appears in the right panel. We call this operation document selection.

pan north-south and east-west to view the page in its entirety. Full-scale layout avoids the serialization-related problems of NSL, but does require of the user a burdensome amount of panning and scrolling, and also requires of the target browser a significant amount of CPU and memory to compute a rendering of the document.

Most importantly, user studies have shown that both rendering modes lead to disorientation; that is, people often lose track of their location in the document while navigating through it [3][8].

2 A Hybrid Solution

CENTAUR renders content in two side-by-side panels, as shown in Fig. 1. The two panels can be displayed using a frameset. Other techniques such as CSS absolute positioning or HTML tables may be used when the browser lacks support for frames.

The left panel of the display is a raster image of the entire document. The document image is partitioned into visually discrete regions. Each region is individually hyperlinked. Regions are visually demarcated; for example, with different background colors or with uniquely-colored borders. Typically, the entire thumbnail view is rendered as an imagemap; however, in the case that the target browser lacks imagemap support, the thumbnail may be rendered using the HTML `<table>` element, with hyperlinked images comprising the table's cells.

The right panel contains a NSL view of the currently-active region of the document. This panel is represented as XHTML markup. Vertical scrolling may, as usual, be required to access the full contents of the region.

2.1 Finding Regions

A key functional component of CENTAUR is partitioning a web page into regions. Although the "quality" of a partitioning is a subjective notion, we can set forth some guiding principles:

- Individual regions should be clearly visible in the thumbnail image. Regions that are extremely small (just a few pixels) in one or more dimensions are undesirable.
- Each region should be contiguous and convex.
- Partitions should follow the principle of least astonishment: paragraphs should not be separated across regions, nor images from their captions, and so on.
- Regions should not exceed a prespecified size.

"Size" may be defined many different ways. A simple but effective definition is this: the *size* $|x|$ of a region x is the number of pixels occupied by the content (text, images, form controls, etc.) in that region.

It has been noted that most published content, print and electronic, follows a grid-based layout, where content streams through rectangular regions on the canvas [2]. Web browsers represent the grid layout of an HTML document using a hierarchical rectangular structure known as a frame tree [9]. Not to be confused with the HTML `<frame>` element, a frame is a visual pane that contains content and the location of that content on the canvas. Frames may contain other frames, in a hierarchical way. Together, this frame tree corresponds to the hierarchy of the underlying XML-based Document Object Model (DOM). For example, a 2x2 HTML table has a frame that corresponds to the table, and four children frames, one for each `<td>`.

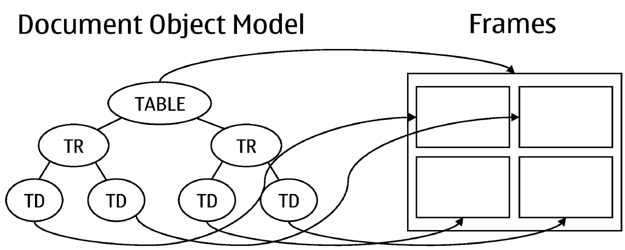


Fig. 3. DOM and corresponding frame tree representation of an HTML excerpt representing a table.

Frames are a good starting point for region-finding, since they capture the hierarchical visual structure of a web page. To narrow the search space for regions, CENTAUR considers only regions that correspond to individual nodes in the frame tree, or to a sequence of sibling nodes. But even with this restriction, the search space is too large; after all, a frame with n children has 2^{n-1} different partitions into regions. CENTAUR therefore employs a greedy search strategy. Working from the root frame node, CENTAUR recursively divides the frame tree, selecting divisions that minimize the variance in size of the resulting regions.

More concretely, imagine that a frame node x has children $\{x_1, x_2, x_3, \ldots x_n\}$. Inserting a seam somewhere in this sequence, the expected size of the resulting regions is $E=(|x_1|+|x_2|+\ldots+|x_n|)/2$. The score S_k of a seam before child k is the inverse of the variance in size of the resulting halves:

$$S_k = \frac{1}{(|x_1|+|x_2|+\ldots+|x_{k-1}|-E)^2 + (|x_k|+|x_{k+1}|+\ldots+|x_n|-E)^2}$$

In other words, a split that results in a more even division of content receives a higher score.

Notice that this region-finding method takes no account of the actual content of a frame node. Ideally, a recursive region-finding algorithm as above would notice that a frame node consists of two child frame nodes with very similar (or dissimilar) characteristics; for example, both consist of a list of hyperlinks, or both are pure text, or both contain only images.

After generating the thumbnail image for the left panel, CENTAUR analyzes the page and selects the "main content" as the region that is initially shown to the user. Recent work has demonstrated automatic techniques for skipping past navigational, branding, and search forms at the top of a web page, and identifying the start of main content [5].

To illustrate, notice that in Fig. 2 the right panel contains content from the "main" portion of the web page: a region containing the day's headlines, as opposed to logos, advertisements, or navigation links.

3 Previous Work

Like CENTAUR, the Thunderhawk HTML browser [1] offers a two-panel viewing mode: one panel is a high-level image of the page, and the other panel contains a zoomed-in view of the active region. The key difference is that Thunderhawk does not itself generate a partitioning of the page; instead, it requires the user to specify a rectangular region of interest within the overview image. It is this region which Thunderhawk renders, zoomed-in, in the second panel.

Also related is Frayling's SmartView system [6]. Within a single panel, SmartView offers either a thumbnail view of the entire page or a NSL view of one region of the page. SmartView contains a region-finding subsystem which relies on a few heuristics involving table and form information. CENTAUR extends the concept of SmartView to a side-by-side view and offers several important enhancements, including region labeling and document analysis to locate the main content region(s).

References

1. BitStream Thunderhawk Browser: http://www.bitstream.com.
2. A. Hurlburt (1977) *Layout: The Design of the Printed Page*. Watson-Guptill Publications. New York.

3. M. Jones, G. Marsden. N. Mohd-Nasir, K. Boone and G. Buchanan (1999) "A site-based outliner for small screen web access." In *Proceedings of the 8th International World Wide Web Conference.* Toronto, CA.
4. HTML 4.01 Specification (1999). Available at http://www.w3.org/TR/html4/.
5. A. Lee and V. Hanson (2003) Enhancing web accessibility. In *Proceedings of the 11th Annual ACM International Conference on Multimedia*.
6. N. Milic-Frayling and R. Sommerer (2002) SmartView: Enhanced document viewer for mobile devices. *Microsoft Technical Report* MSR-TR-2002-114.
7. Opera's Small-Screen Rendering: http://www.opera.com/products/smartphone/smallscreen/.
8. V. Roto and A. Kaikkonen (2003) Perception of Narrow Web Pages on a Mobile Phone. *Proceedings of the 19th International Symposium on Human Factors in Telecommunications.* Berlin, Germany.
9. C. Waterson (2002) Introduction to Layout in Mozilla. Slides available at http://www.mozilla.org/newlayout/doc.

Smartphone Views: Building Multi-device Distributed User Interfaces

Micael Sjölund, Anders Larsson, and Erik Berglund

Department of Computer and Information Science
Linköping University
{x03micsj,andla,eribe}@ida.liu.se

Abstract. This paper introduces a prototype of a *distributed user interface* (DUI) on dual devices, a workstation and a Windows Mobile-powered smartphone. By porting the XML-compliant GUI system Views to the smartphone platform, we explore one possibility of distributing GUI components among heterogeneous devices. We describe problems and conclusions from designing and implementing the system.

Keywords: Distributed user interfaces, ubiquitous computing, XML GUI description, mobile computing, adaptive user interfaces, smartphone, Views

1 Introduction

In a world of increasing number of computational devices with graphical display capabilities in our immediate surroundings we can efficiently distribute user interfaces among devices to create working units. By taking advantage of the individual benefits of each device in the environment (e.g. the portability of the handheld, the display capabilities of the desktop computer), a whole new methodology of computer interaction becomes possible.

Around us, numerous heterogeneous devices with advanced display capabilities, provide a platform for flexible and natural interactivity for mobile work. One of these possibilities are distributed user interfaces (DUI) [1], where GUI components are spread among devices. Several projects have explored how to automatically generate user interfaces for heterogenous devices [2,3,4,5,6,7]. Numerous device-independent GUI specification languages has been reviewed in [8].

In our vision, an application can not only display a version of its user interface on other devices but *move parts* of it when other devices are proximate, and doing so without manual configuration on the device it distributes it to. In this distributed user interface, the devices displaying the interface are not all equal. The logic of the application lies only on one device, but the interaction components are distributed. Rather than moving an entire interface to a small device, with the proper UI transformation performed, we aim for casual distribution of GUI elements from applications directly.

This paper reports on an initial implementation application for the development of a GUI distribution framework to mobile phones, or more specifically the

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 507–511, 2004.

Windows Mobile platform and the smartphone. Here we describe how we port the XML-compliant GUI system Views [9], see Section 2, to the Windows Mobile platform as a distribution engine for DUIs. We also describe the implementation of a simple DUI example, a distributed remote control GUI to Windows Media Player.

2 Porting Views to the Smartphone

The implementation of our DUI system using the smartphone as a platform for GUI distributions requires a general-purpose way of describing GUI layout, without the need of compiling the interface in advance.

Views [9] is an XML-compliant GUI description system for C#. It is a platform independent description language, but it's main implementation is for the .NET Framework platform, using Windows Forms to build the GUI controls.

The choice of Views as GUI description language was based on several considerations. Its implementation for the .NET framework is based on Windows Forms, and the transition to the smartphone would be fairly easy. Using Views on the smartphone provides the controls and functionality we need to study different aspects of distributed user interfaces.

Getting a Views system up and running on the smartphone was not a big problem itself, but it came with some limitations. We describe some of the issues and how we decided to approach them.

Control set. Since both the desktop Views system and the Windows forms for smartphone only handles their own subset of the available controls for desktop Windows Forms, the initially available controlset in Views for smartphone consisted of an intersection of the two sets. Although small, the controlset remaining initially allows us to use event-generating controls, as well as input and feedback controls.

GUI events. Events from GUI controls in Views are dispatched as a string with the name of the control that generated the event. This way of generating events is well suited for building a DUI, since they are easily distributed over network. Events that are generated at a distributed part of a UI is simply sent back to the device which handles the application logic. However, there were difficulties in porting the event system to the smartphone. The event string is stored in the `Control.Name` property of the controls, and extracted when an event occurs. Unfortunately, the .NET Compact Framework version of Windows Forms does not have this property implemented, so we had to construct wrapper classes for the controls that generates events, with a `Name` property implemented. An example of this is our `DUICheckBox` that inherits the `System.Windows.Forms.CheckBox`.

Input actions. In Windows Forms for the smartphone, GUI buttons are not in the available control set. Input for The Windows Mobile-powered smartphone are mainly performed through the two softkey buttons, which are controlled from the two phone keys directly below them. To be able to take

Fig. 1. The DUI Views remote control application on the smartphone

Fig. 2. The Windows Media Player embedded in a DUI application

direct user inputs with the lack of GUI buttons, we implemented support for softkey menus. Since pull-down menus was not supported by desktop Views, this had to be done from scratch. For this prototype, we mapped the left softkey button to *Done* (quit), and the right softkey button to a user-defined menu system. The menu can be implemented in the GUI description file with the rest of the GUI, with nested submenus and menuitems. The menuitems are named and generates events like any other event-generating Views widgets.

Data input. Data input (text fields, checkbox state) on the smartphone raises the question of when and what to send back to the DUI server, since no application logic exists on the distributed parts of the GUI. We have solved it by sending the state of all controls back to the application when an event is raised in the gui.

Positioning. For control positioning, the desktop Views implementation offers commands for both horizontal and vertical alignments. We have chosen to support only vertical alignment for the smartphone. This is both simple and conforms well with the Windows Mobile UI guidelines [10].

Feedback. Feedback on Views controls are performed with specific methods, such as `PutText()` to set the text displayed in a text field. Communicating feedback over the network is as easy as event communication, since the controls are referenced by their names.

3 Implementing a Media Player DUI Remote

As an initial application, a DUI remote controls for the Media Player has been developed. Using a smartphone with the ported version of Views, it has been possible to spread the GUI of the controls from the workstation and thereby distribute screen space. Fig. 1 and Fig. 2 shows screenshots of the implementation.

There are numerous implementations of remote control systems for PDA and smartphone, both research and commercial [11,12,13,14]. However, none of them has the DUI angle of study, with the casual distribution of GUI components we aim for. Our remote control serves as an example of a more general approach of studying GUI distribution.

We are currently using sockets for network communication to share events and feedback between the devices. By using ActiveSync to connect with sockets, this allows for both wireless connection to the smartphone via bluetooth, and wired USB connection.

4 Future Work

The work so far has raised several questions about for instance on how to communicate events and feedback between the devices. The continuation of the project will look deeper into these issues. Also, we want to implement support for more GUI controls in smartphone Views, such as better feedback and multi-choice components.

5 Conclusion

In this paper we show how we have ported Views to the Windows Mobile platform for the smartphone and a sample application in the form of a DUI remote control for Windows Media Player. These implementations represents steps towards creating general-purpose distributable GUIs.

In order to support the development of distributed user interfaces, we need to develop platforms and development models for the construction of GUIs with distributable parts. Porting Views to the Smartphone was not straight forwards since it came with some limitations described in Section 2. However, we have found ways to overcome the differences and to provide a process to directly distribute GUI parts from applications to other slave applications.

Acknowledgements. This project is possible due to funding and sponsoring from Santa Anna Research Institute and Microsoft Research.

References

1. Larsson, A., Berglund, E.: Programming ubiquitous software applications: Requirement for distributed user interface. (2004)
2. Nichols, J., Myers, B.A., Higgins, M., Hughes, J., Harris, T.K., Rosenfeld, R., Pignol, M.: Generating remote control interfaces for complex appliances. In: Proceedings of the 15th annual ACM symposium on User interface software and technology, ACM Press (2002) 161–170
3. Nichols, J., Myers, B.A., Litwack, K.: Improving automatic interface generation with smart templates. In: Proceedings of the 9th international conference on Intelligent user interface, ACM Press (2004) 286–288
4. Gajos, K., Weld, D.S.: Supple: automatically generating user interfaces. In: Proceedings of the 9th international conference on Intelligent user interface, ACM Press (2004) 93–100

5. Ponnekanti, S., Lee, B., Fox, A., Hanrahan, P., Winograd, T.: Icrafter: A service framework for ubiquitous computing environments. In: Proceedings of the 3rd international conference on Ubiquitous Computing, Springer-Verlag (2001) 56–75
6. Han, R., Perret, V., Naghshineh, M.: Websplitter: a unified xml framework for multi-device collaborative web browsing. In: Proceedings of the 2000 ACM conference on Computer supported cooperative work, ACM Press (2000) 221–230
7. Myers, B.A.: Using handhelds and pcs together. Commun. ACM **44** (2001) 34–41
8. Nylander, S.: Approaches to achieving device independent services - an overview. Technical Report T2003-16, SICS (2003)
9. Bishop, J., Horspool, N.: Developing principles of gui programming using views. In: Proceedings of the 35th SIGCSE technical symposium on Computer science education, ACM Press (2004) 373–377
10. Webster, C.: Crafting smartphone user interfaces using .net compact framework (2003) accessed March 31 2004.
11. Sanguinetti, A., Haga, H., Funakoshi, A., Yoshida, A., Matsumoto, C.: Frecon: a fluid remote controller for a freely connected world in a ubiquitous environment. Personal Ubiquitous Comput. **7** (2003) 163–168
12. RuttenSoft: Media remote 2004. (`http://www.ruttensoft.com/smartphone/mediaremote.htm`) accessed March 31 2004.
13. Salling Software: Salling clicker. (`http://homepage.mac.com/jonassalling/Shareware/Clicker/`) accessed March 31 2004.
14. Feldbusch, F., Paar, A., Odendahl, M., Ivanov, I.: The btrc bluetooth remote control system. Personal Ubiquitous Comput. **7** (2003) 102–112

The Lack of Intelligent Design in Mobile Phones

Georg Strom

Department of Computing Science, University of Copenhagen,
Universitetsparken 1, DK-2100 Cph. O., Denmark
georg@diku.dk

Abstract. This paper describes how the organization of the development of mobile phones may reduce usability and user satisfaction, and it suggests possible improvements to the development process. It compares the evolution of mobile phones with the evolution of animals and plants, and it describes how the evolution of mobile phones is driven by the competition between specific characteristics of phones; similar to the way organic evolution is driven by the competition between genes. It describes how characteristics of mobile phones compete both in markets and in development organizations, and criteria that determine which characteristics survive and spread in development organizations.

1 Introduction

Mobile phones are much less exciting than they could be, they are complex and difficult to operate, and there is a lack of phones for common groups with special needs, for instance people wearing gloves while working or doing sport. This was stressed repeatedly during the special area discussion of mobile phones at the CHI2004. It appears that most mobile phones today look like plastic pebbles or toys for teenage girls; they are becoming increasingly similar and difficult to operate.

In contrast, I have on and off been involved with mobile phones for eighteen years. I know from my own experience that most people developing mobile phones are bright, dedicated and focused on producing what they believe are the best phones possible. Therefore, it is interesting to discuss why they do not appear to be more successful.

2 Design Versus Evolution

George Basalla [1] states in his knowledgeable book on the evolution of technology that, in contrast to the evolution of life, the development of technical artifacts is "the result of a conscious process in which human judgement and taste are exercised...". This is what most technological companies want to believe about their own development processes, and it is similar to what some Christian groups state about the evolution of plants and animals [7]: That it is governed by some sort of intelligent design (which has given this paper its title).

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 512–516, 2004.

In reality, there is a similarity between the evolution of mobile phones and the evolution of plants and animals. In the organic evolution the competition is not between single organisms or even species, but between genes and thereby characteristics [3]. Similarly, within the evolution of mobile phones it is not specific models or brands that compete, but characteristics: Features, specific input devices or specific ways to group or organize functions in a phone. Successful characteristics may even be considered as what biologist Richard Dawkins describes as memes [3]: Ideas that spread through part of human society in the same manner as successful genes spread through a population of plants or animals.

3 Survival in Markets

Mobile phones fulfill both practical and strong emotional needs [5], and they fit all the characteristics that Rogers [6] describes for innovations that spread rapidly:

- Mobile phones offer a large relative advantage compared to previous solutions as for instance payphones.
- Making a phone call is compatible with existing values.
- It is easy to use a mobile phone for making a call.
- It is possible to observe the advantage other people gain from using it.
- It requires only a small investment to use on a trial basis (in particular when service providers subsidize the acquisition).

This means that the characteristics shared by all mobile phones are so superior that almost any phone can succeed in an early market.

In contrast to other consumer electronics, as for instance cameras and music systems, the quality of the basic services (voice and SMS communication) are exactly the same in all mobile phones. In addition, the basic services of all mobile phones require RF-circuits and a considerable amount of processing power, making it unfeasible to offer a really low cost phone with only the most basic functions. This means that compared to other consumer electronics the tiering between low-cost and so-called professional mobile phones is less visible and more artificial.

At the same time, it is difficult for users to evaluate the usability of a phone or the usefulness of specific non-basic functions before they have bought the phone and used it for a period of time. Users are therefore forced to base their purchasing decision on other criteria.

A significant group of users believe that the characteristics of a particular brand of mobile phones are superior, even though all brands and all models of mobile phones consist of a mixture of good and bad characteristics. Other users may buy a phone with inferior characteristics, because it also contains a highly advertised game or a function for composing ringing tones. Finally, users who find it difficult to evaluate the functionality of different phones may buy the cheapest phone that looks acceptable. The price is the most visible characteristic, and the one that is most easy to compare. This means that mobile phones with some inferior characteristics often are successful even in mature markets.

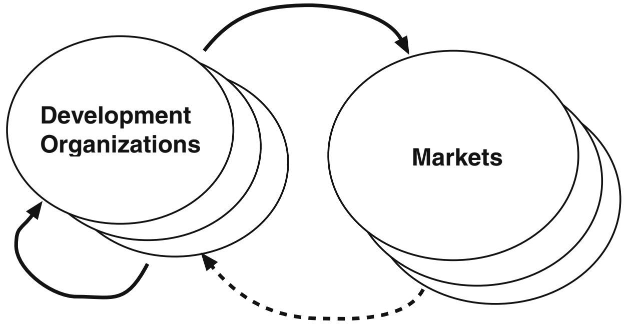


Fig. 1. The relationships between development organizations and markets.

4 Relations Between Markets and Development Organizations

The attractiveness of a specific model of a mobile phone depends on branding and advertising and on the total attractiveness of all its characteristics. Even when users buy or reject a particular model because of one specific characteristic, it is in most cases impossible to deduce this from sales figures. This means that the advantages of different characteristics are communicated only through media and market surveys.

Most people doing and communicating field studies are university graduates with middle class values, and they tend to select and filter information from markets such that it fits their own values and background. They may find it difficult to capture that an elderly user fears and feels ashamed of possible memory impairments. If a young user describes sexually explicit usage and characteristics, they will probably not explore these aspects in their report of the field study.

Both field studies and sociological studies of the use of mobile phones are too large to use directly as input in a development process, and it is necessary to translate them into information about characteristics of mobile phones. That is difficult and time consuming. It is far easier for developers to look on characteristics of other mobile phones and to be inspired by them, and it is a well-established practice to buy and study competing products. My own experience indicates that information about characteristics spread more easily between development organizations in different companies than between markets and development organizations. See figure 1. (Different characteristics actually behave as memes: Ideas that spread through and between development organizations.)

5 Competition in the Development Organization

The different characteristics compete directly with each other when people in the development organization decide which characteristics to include in a new model. A characteristic may be included because it usually is included, and because nobody questions it, or when someone with influence in the development organization pro-

motes and defends it. He or she may do so because the characteristic seems useful or for a number of other reasons.

Basalla [1] describes how the creation of technical innovations often is more similar to play or fantasies about what might be possible, than to the solving of already recognized problems. A characteristic that excites developers, for instance by including a new solution of a difficult technical problem, may survive and be included in new products, even though its usefulness is limited.

Basalla [1] describes how designers tend to accommodate new materials to old forms, to be skeumorphic. One example is the tabs used in some mobile phones that has been taken over from Microsoft Windows and before then from paper based folders. Skeumorphic characteristics are easier to communicate and therefore tend to spread and survive in a development organization. (Even when they offer little or no benefit to users.)

Tracy Kidder [4] describes how the structure of a product tends to reflect the structure of the organization that has created it. It is difficult to combine functions across organizational boundaries and trying to do so may result in power struggles. This means that mobile phones developed by large and complex organizations tend to have a large and complex structure. It also means that a function, a characteristic, more easily survives when it fits the organizational structure, for instance when a large group already has been dedicated to developing it.

To suggest that a specific characteristic shall be removed will likely cause a conflict with its proponents, and in most cases they are able to argue that there are situations in which it may offer some benefit. The consequence is that the complexity of mobile phones in general increases over time, even when it reduces their usefulness.

6 Conclusion

Despite its imperfections, evolution has created much better mobile phones than what was possible through any sort of intelligent design. Evolution works in parallel, it goes through far more different combinations of characteristics than a single designer can do, and it has created phones that nobody could imagine when the standards of the GSM system were defined in the eighties.

What about the problems described in the introduction?

The lack of excitement in mobile phones may be more of a problem for developers of mobile phones than for users, or the problem may be that the phones are not as exciting for users as for developers. The developers are primarily excited by what they have created during the development of the phones, whereas users are more excited by what they can express and do with their phones.

There is a lack of ruggedized phones for use at building sites or during sailing and skiing, in fact there is a lack of phones that is rugged enough for everyday use. However it is likely that only few phones with such characteristics will evolve: Their price, their most visible characteristic, will be so high that it is difficult for them to compete in most markets.

In the Philippines I have seen how mobile phones often are the only accessible means of communication, and how they together with SMS services are used as an alternative to e-mail and the Internet. The situation is probably the same in rural areas in many other emerging countries, and there is a need for phones that are suitable for such extended use and able to withstand a humid and dusty environment. The superior characteristics of such phones make it likely that some designers will promote and defend them. However, it is likely that users in such areas will perceive them as less attractive or fashionable than ordinary phones, and that the phones will be too expensive to reach the target group without subsidies (similar to the Freeplay wind-up radios and flashlights).

The effort to provide wireless access to train schedules, news and other information services has failed miserably. Not least because the user interfaces of the phones have been too difficult to operate. Other services may suffer the same fate as long as the characteristics of mobile phones mainly are determined by internal factors in development organizations.

The design of more usable phones requires organizational changes. Bergman, and Haitani [2] indicate that the development of a less complex and more usable product requires that the design of the entire user interface is done by a small group where it is easy to make decisions, and they stress that user involvement was a prerequisite for making the first PalmPilot: A simple and successful product. A first step may be the use of participatory design involving a variety of users: Children, teenagers, elderly people, members of subcultures and others. The second step is to accept the thoughts and experiences of the users, in order to design mobile phones that make it possible for them to express themselves and fulfil their own goals, even when the designers do not agree with them.

References

1. Basalla, George: The evolution of technology. Cambridge University Press, UK Cambridge (1988)
2. Bergman, Eric & Rob Haitani: Designing the PalmPilot: A conversation with Rob Haitani, in Information Appliances and beyond, Eric Bergman (ed.). Morgan Kaufman, San Francisco (2000)
3. Dawkins, Richard: The selfish gene. Paladin, London (1978)
4. Kidder, Tracy: The soul of a new machine. Penguin Books, Great Britain (1982)
5. Ling, Rich: The Mobile Connection: The Cell Phone's Impact on Society. Elsevier Inc. (2004)
6. Rogers. Everett M., Diffusion of innovations, 3rd ed.. New York Free Press (1983)
7. Shermer, Michael: Why people believe in weird things. Henry Holt and Company LLC, New York (2002)

Visual and Interaction Design for 3G Mobile Phone Interfaces

Anders Norman, Sofia Svanteson, and Jesper Wiking

Ocean Observations
Åsögatan 108
118 29 Stockholm, Sweden
{nome, sofia, jesper}@oceanobservations.com

Abstract. This is a full day tutorial regarding visual and interaction design for 3G mobile phones. The first part of the day will be spent discussing the total user experience, user modes, information architecture, visual concepts and visual graphics such as iconography, widget design, fonts, colors and page design. The second part of the day will be spent on an exercise where the participants will design their own GUI concepts in groups.

1 Introduction and Motivation

If you know how to use the empowering mobile features such as color, complex graphics and animation, you can design attractive, usable and targeted graphic user interfaces for mobile phones. Although the screens of today enable us to use these features to create designs and products based on peoples needs, interests and cultural values we see a slow progress in this area. A mobile phone's content and GUI design is often the very same weather the phone is marketed towards a person in a business mode where a superior calendar is his or her lifeline, or a teenage skate-punk whose main interest might be finding the nearest music venue and let the mates know.

When manufacturers talk about phone design, they most often refer to the design of the hardware. GUI design is seldom considered when marketing a new phone, which is surprising since the GUI is a big part of the experience of the phone and thus is a vital part of the relationship between a brand and its user.

2 Learning Objectives

This tutorial teaches you, through lecturing and easily understood break-out sessions, why a focus on attractive and usable GUIs is important and how to design them.

The tutorial starts with a crash course in User Centered Design where we answer questions such as; What is user centred design, why is it important and what kind of values can it add to both users and stakeholders (manufacturers, operators, 3rd parties)? We will also show the impact a user centric design process have on a Brand's ability to build strong relationships towards users.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 517–518, 2004.

Then the tutorial moves on with an introduction to cultural aspects – how are needs, values and motives different between and within countries and continents? We will give examples of how to structure and design information for different user modes and profiles.

The participants will also learn how to design efficient menus and navigation, the importance of semiotics; ensuring icons, labels and headlines help user identify what they are looking for, concept design, how to make widgets, fonts, colors and shapes come together into a clear and visually pleasing interface, and finally how to balance and use animated graphic elements.

The tutorial will also include and exercise and real case studies in order to show and discuss how GUIs of the same phone model can look different when targeted towards different user modes/profiles.

3 Intended Audience

The tutorial is suited for interaction designer, visual designers, usability specialists, developers, marketing directors, decision makers or anyone with an interest in mobile phone GUIs. The tutorial does not require any knowledge and as such anyone with a basic understanding of the features and functions of general usage of mobile phones can attend.

Handheld Usability: Design, Prototyping, and Usability Testing for Mobile Phones, PDAs, and Pagers

Scott Weiss

Principal, Usable Products Company, sweiss@usableproducts.com

Abstract. This tutorial presents handheld device UI design, prototyping, and usability strategies. Handheld devices include pagers, PDAs, and mobile telephone handsets. Hardware UI, operating environments, and wireless networking will be presented. Information architecture, paper prototyping, and usability testing of handheld devices will be taught with team exercises.

1 Description

Handheld devices have problematic input mechanisms and tiny displays. The devices do not support layered windows well, and drag-and-drop does not exist in their interaction paradigms. Designers must understand the types of input afforded by each device and operating environment, as well as their display characteristics.

Handheld devices are used by people on the go. Attention spans are limited, as the devices are brought into situations where they are secondary to the user's focus. Desktop computers receive dedicated focus, but handheld devices are given only fragmented bits of attention.

Network connectivity for handheld devices is slow, unreliable, and expensive. In this tutorial, designers and usability experts will benefit from understanding the basics for how these networks operate and how to design effective user interfaces to cope with the limitations. Usability testing of handheld devices, given these constraints, is equally challenging.

A brief introduction to types of handheld devices and operating environments sets the stage for more in-depth discussion and hands-on learning activities centered on design & information architecture, paper prototyping, and user testing. Exercises focus on applying the skills and strategies learned to design a user interface for a mobile telephone handset.

1.1 Background of the Handheld Device Platform

This section starts with a brief overview of computing and telecommunications convergence, with historic examples. It surveys handheld environments, use cases, device hardware, and software user interface elements. This section concludes with an introduction to handheld device data, including Java and WAP.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 519–520, 2004.

1.2 Information Architecture in the Handheld Medium

This section discusses how generally used IA processes can be applied to the small display. Audience definition, task analysis, and information architecture will be presented. Site/Application structures will then be discussed, along with methods of documentation, issues of nomenclature, and navigation structures. The first exercise is a brainstorming session to identify user interface applications for handheld devices. Attendees will select a UI to design and prototype, and be assigned to teams.

1.3 Online and Paper Prototyping

This section starts with a discussion detailing major considerations of prototyping a handheld application, including device differences, a survey of software tools available, and the pros and cons of different strategies. Based on the site architectures developed in the previous section, the exercises in this section are aimed towards the development of a richly detailed prototype. The exercises build skill and confidence in paper prototyping, and provide material to be used in the final tutorial section.

1.4 User Testing of Handheld Devices

Evaluation of UI designs is presented in this section. A user testing lab setup for handheld devices is discussed, including developing a participant screener, defining tasks, data recording, and room/equipment setup. Variations of this structure are explored. Quick usability tests of two of the prototyped user interfaces will be conducted to demonstrate handheld device usability interview strategies. Issues regarding the interaction are noted, and possible solutions are devised.

2 Presenter

Scott Weiss is the Principal of Usable Products Company, an ease-of-use consultancy focused on mobile device usability and design. Usable Products' clients include Vodafone, Sprint PCS, LG Electronics, Microsoft, and Intel. Scott's work experience includes career positions at Apple, Microsoft, Sybase, and Autodesk. He is the author of *Handheld Usability* (Wiley: 2002), which covers design, prototyping, and usability for mobile device applications. Conferences\

3rd International Workshop on: "HCI in Mobile Guides"

Keith Cheverst[1] and Barbara Schmidt-Belz[2]

[1] Department of Computing,
Lancaster University,
Lancaster, LA1 4YR.
kc@comp.lancs.ac.uk
[2] Fraunhofer FIT,
Germany.
Barbara.Schmidt-Belz@fit.fraunhofer.de

Abstract. In recent years, we have witnessed a proliferation in research related to mobile guide systems. This is in part due to the increasing availability and affordability of the required enabling technologies and a growing acceptance of the potential for mobile guides in the market place. However, the complex issues surrounding the human factors issues associated with the use of mobile guides still requires considerable investigation if the 1st generation of marketed mobile guide systems are to be usable and offer real value to the user. In this third workshop on 'HCI in mobile guides', the key aim is to provide a vehicle to enable researchers and practitioners to continue to share their understanding and findings relating to HCI with mobile guides.

1 Motivation

Today's mobile user demands easy access to relevant information services from a variety of devices (both personal and situated/public), whenever and wherever they need them. Example applications for mobile guides include: mobile tourism services, indoor and outdoor museum/exhibition/event guides and context-aware directory services. Although the latest mobile devices and information services offer new and enhanced ways to support nomadic users, they also raise challenges concerning interaction modalities, usability, accessibility and trustworthiness.

2 Topics of Interest

A range of topics are relevant to a discussion on the human computer interaction issues relating to mobile guide systems. In this workshop, the following topics are of particular relevance:

- Accessibility for particular groups, e.g. older users, visually impaired etc.
- Approaches to (and results of) requirements capture for mobile tourism.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 521-522, 2004.

- Suitability of different interaction metaphors, e.g. anthropomorphic approaches that cope with the limitations imposed by mobile devices.
- Visualization of the spatial environment, Augmented Reality, 2D/3D maps etc.
- Implications for adaptive behaviour, e.g. location awareness.
- Handling and conveying dynamic information, e.g. changes to available services.
- Leisure/entertainment use (e.g. by games on treasure-hunts or to support spontaneous social gatherings).
- Techniques to facilitate access to heterogeneous and/or distributed services.
- Support for both traditional and social navigation, e.g. supporting anonymous recommendations etc.
- Personalization of services, e.g. use of user modelling techniques.
- Techniques for and experience of user evaluation of mobile guides.
- Novel infrastructures, such as agent-based technology, and their implications for interaction.
- Fault tolerance, trustworthiness, and security.
- Information retrieval and display whilst faced with changing infrastructure conditions.
- Design solutions for "baby interfaces", i,e, small buttons, small screens and small interaction devices (tiny joysticks and tiny pens).
- Issues arising from the opportunities and challenges provided by multimodal user interfaces
- Designing for the wild: new and innovative methods that explore the design of mobile guides in the wild.

This workshop is the 3rd in a series. The first workshop in the series was, however, particularly focused on "HCI in mobile Tourism" [1] while the second workshop [2] was concerned with HCI issues relating to mobile guide systems in general.

3 Program Committee

Lynne Baillie (FTW, Vienna),
Keith Cheverst (Lancaster University, U.K.),
Fabian Hermann (Fraunhofer IAO, Germany)
Eija Kaasinen (VTT Information Technology, Finland)
Chris Kray (Lancaster University, U.K.),
Elke-Maria Melchior (ACIT, Germany),
Stefan Poslad (Queen Mary University of London, U.K.),
Barbara Schmidt-Belz (Fraunhofer FIT, Germany).

References

1. Schmidt-Belz, B. and K. Cheverst. Proc. of 1st Workshop on Mobile Tourism Support, B. Schmidt-Belz and K. Cheverst (Eds), ISBN 3-88457-980-0, August 2002.
2. Schmidt-Belz, B. and K. Cheverst. Proc. 2nd Intl Workshop on Mobile Tourism Support, B. Schmidt-Belz and K. Cheverst (Eds), Udine, Italy, Sept. 2003: www.mguides.info.

Interaction Design for CSCL in Ubiquitous Computing

Lucia Terrenghi[1], Carla Valle[1], and Giorgio De Michelis[2]

[1] Fraunhofer FIT, Institute for Applied Information Technology, Schloss Birlinghoven, 53754 Sankt Augustin, Germany.
{lucia.terrenghi, carla.valle}@fit.fraunhofer.de
http://www.fit.fraunhofer.de

[2] DISCo, University of Milano – Bicocca, Via Bicocca degli Arcimboldi , 820126 Milano, Italy
gdemich@disco.unimib.it, http://www.unimib.it

Abstract. This workshop aims to share interdisciplinary experiences and perspectives in the field of CSCL in the area of Ubiquitous Computing. In particular, interaction and interface design for distributed education; new interaction patterns within and among distributed learning communities: learning experience evaluation; accessibility of learning material in different contexts are the main focus of discussion.

1 Topic

Ubiquitous computing has been affecting our everyday lives by providing us with mixed places in which the *virtuality* of computer-readable data is brought into the physical world. This creates a great opportunity to use technology in appropriate ways to enhance and augment the learning activity in different aspects: by enabling people to interact and collaborate remotely; to enlarge teaching and learning possibilities; to increase students' access to learning opportunities and educational material; by supporting hands-on experiences and situated learning; by favouring a continuous exchange of experiences and perspectives among the members of educational communities; by connecting different learning communities, thus enhancing knowledge building and sharing.

These goals raise new challenges in terms of interaction design, suggesting the need of new forms of interaction patterns between users and environments, and between different groups of users. Design can play a key role in shaping new ways of collaborative learning and knowledge management, and enhance the natural evolutions of learners' sense of place and time towards the experience of living in a mixed reality, in which physical and virtual spaces are blending together, and social relationships become fluid and distributed.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 523–524, 2004.

2 Goals

This workshop aims to raise discussions on the topic, and leverage a share of experiences among people addressing these aspects from different perspectives. In this sense we foster the exchange and interaction among participants from different communities, such as interaction design, education, CSCL and CSCW, software engineering, ethnography and sociology, enhancing an interdisciplinary approach and cross-fertilization among communities. The workshop will then be the opportunity for participants to look at the topic from multiple points of view, to share and review their research agenda, and to make contacts that can enrich their work.

3 Content

Themes that are relevant for this workshop include, but are not limited to:

- *interaction and interface design for distributed education*: interface design for multi-platform applications and mobile learning; shared interfaces design; tangible and multimodal interfaces for education; interface design for social awareness and collaboration;
- *new interaction patterns within and among distributed learning communities*: recognition of collaborative, educational and social patterns; activity theory approaches; ethnographic studies; comparative studies on different domains and device technology;
- *learning experience evaluation*: usability and pedagogical evaluation techniques; user experience evaluation methods and approaches;
- *accessibility of learning material in different contexts*: storage and retrieval of Reusable Learning Objects into digital libraries that are accessible from different devices; context-adaptive content annotation and presentation; educational content generation and visualization on mobile devices; knowledge building and management through distributed learning communities.

4 Organizers

Lucia Terrenghi works as HCI researcher and user interface designer at the Fraunhofer FIT. Here she works at the design and evaluation of applications for ubiquitous computing and e-learning. **Carla Valle** works as CSCW researcher at the Fraunhofer FIT. Her research addresses the areas of CSCW, Mobile Computing and Decision Support Systems. Prof. Dr. **Giorgio De Michelis** is the Director of the Department of "Informatica, Sistemistica e Comunicazione" and teaches "Theoretical Computer Science and Information Systems" at the University of Milan - Bicocca. He carries out research on models of Petri Nets, Computer Supported Cooperative Work, where he develops prototypes of support systems for cooperative processes, Communityware and related topics.

Mobile HCI 2004 Workshop on Location Systems Privacy and Control

Ian Smith[1], Giovanni Iachello[2], and Mika Raento[3]

[1] Intel Seattle Research, 1100 NE 45th Street, 6th Floor,
Seattle, WA 98105, +1 (206) 633-6555
Ian.e.smith@intel.com

[2] Georgia Institute Of Technology, 801 Atlantic Drive
Atlanta GA, 30332-0280 +1 (404) 385-0257
giac@cc.gatech.edu

[3] Dept. Of Computer Science, Univ. Of Helsinki,
Room D419, P.O. Box 26, Teollisuuskatu 23
mika.raento@cs.helsinki.fi

1 Introduction

People are increasingly carrying location-aware devices (*i.e.*, able to determine their own location, and therefore that of the user, in physical space). A variety of such location systems are currently deployed or under development, from the global mobile telephony infrastructure [4] to schemes based on infrared badges, Bluetooth, GPS, or WiFi (802.11).

These location systems raise several concerns, among which privacy, security and information control are at the forefront of social and legal discourse (*e.g.*, [1, 2]). Some countries have adopted pertinent, if partial, legislation. Other countries' regulatory regimes lag behind. Operators and service providers are uncertain of the legal context. Users are unaware of their options, abilities and rights. From the technical perspective, while much work has been done in the privacy and ubiquitous computing communities, it rarely influences how new systems are designed and how technology is introduced within existing social and organizational structures.

Within the scope of the Mobile HCI 2004 conference, we are interested in addressing privacy and information control issues from the user's perspective. We intend to approach these questions from a multidisciplinary, human-centered approach, integrating an analysis the technical characteristics of location systems with relevant usability, social and legal considerations. We hope that by addressing different concerns (*e.g.*, personal privacy, data protection, system integrity, cost factors) we will be able to refine the current discussion in the field, by identifying and characterizing salient issues, and proposing a range of adequate protection tools for each.

We are particularly interested in the following issues:

– **Understanding.** Do users understand how the system works and what they are disclosing to the location system?
– **Cost/benefit analysis.** What benefits do users gain from disclosing their location information? How do users effectively assess those benefits?
– **Privacy Enhancing Technologies.** How can technology be used to prevent the disclosure of information that the user desires to be kept private?

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 525–526, 2004.

- **Legislation.** In what ways may disclosed information be used in different regulatory regimes? How should technology be parameterized to satisfy these requirements? How should legislation adapt to privacy-preserving location systems?
- **Culture.** How do social conventions and expectations vary across cultures?
- **Social dynamics.** In what ways is the disclosure of a user's location to a service provider, individual or organization similar to or different from other disclosures people make in everyday life?
- **Trust.** What organizations or individuals do users trust with their location information and why?

2 Intended Audience

This workshop intends to stimulate a discussion which takes into consideration the three stances, which previous research [3] on privacy on the Internet has shown to be mostly representative of social trends: of those who would like to push technology and figure out privacy issues later, of the deeply worried privacy advocates, and of the those without a clear opinion on the topic. We would like to attract a mix of academic researchers, telecom operators, developers and policymakers in order to spur a comprehensive discussion of the global consequences of location systems. By thinking through and understanding diverse perspectives, the organizers think that they can move forward with the development and use of location systems and yet do so in a way that is respectful of the privacy needs and desires of users.

3 Call for Contribution

The organizers encourage people with an interest in the questions outlined above to participate in this workshop, by presenting a position paper, legal analysis, or user study. We are soliciting original contributions, (1-4 pages) on the following topics:

- regulatory issues: need and scope for novel legislation in the field;
- social issues: user studies and the effect of social differences on design issues;
- usability issues: how to build applications that enhance user's understanding of the underlying principles and functionality;
- architectural issues: how to compromise between privacy, security and market needs in a multilateral perspective.

References

[1] Ackerman, L., et al., Wireless Location Privacy: Law and Policy in the U.S., EU and Japan, The Internet Society, www.isoc.org
[2] Batista, E., U.S. Mobilizing Phone Assault, Wired News, August 21, 2000.
[3] David B., *et al.*, Tit for Tat in Cyberspace: Consumer and Web Site Reponses to Anarchy in the Market for Personal Information, UNC Journal of Law and Technology (4) 2, p.217.
[4] Federal Communications Commission, Enhanced 911, http://www.fcc.gov/911/enhanced/

Mobile HCI and Sound

Simon Holland[1], Robert Day[1], Grégory Leplâtre[2], and Alistair Edwards[3]

[1] Department of Computing, The Open University,
Milton Keynes, UK
{s.Holland, r.h.day}@open.ac.uk

[2] School of Computing, Napier University,
Edinburgh, UK
g.leplatre@napier.ac.uk

[3] Department of Computer Science, University of York,
York, UK
alistair@cs.york.ac.uk

1 Introduction

Sound plays an increasingly varied and vital role in mobile and ubiquitous user interaction. One reason is the limited screen real-estate available in typical mobile devices. Another reason is that many mobile devices are used in *minimal attention situations* These are situations in which the user has only limited attention available for the interface: the user's eyes may be busy elsewhere; and the user may be busy avoiding the normal hazards of moving around, and engaging with real-world tasks. In many circumstances, such interactions will involve non-speech audio and gesture to afford natural means of access to information, to other people, and to services and situations in the environment.

A complementary set of issues concerns Mobile HCI and Music. Mobile interaction technologies will create new opportunities for musicians to compose and perform music on the move and away from studios and concert halls, including in new shared ways. The new technology will also create opportunities for listeners; for new approaches to music education, and for the spontaneous sharing of music by all**.** The workshop will actively seek to benefit from cross fertilisation between researchers involved in both areas. The workshop will seek to gather researchers from a variety of disciplines, to examine challenges, issues, problems, principles, findings new applications and demonstrations related to Mobile HCI and Sound.

2 Topics

Topics of interest will include, but not be limited to:

- Sound as an interaction element in minimal attention interfaces
- Spatial sound in mobile HCI
- Auditory displays in mobile & ubiquitous user interaction
- Sound and gesture in mobile interface techniques

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 527–528, 2004.

- Sound for sighted mobile users
- Sound for mobile users with visual disabilities
- Sound in social navigation
- Complementarity between non-speech and speech audio
- Complementarity between audio and haptic mobile interfaces
- Mobile, Ubiquitous and Tangible HCI in music performance
- The Mobile Musician Machine Interface
- Innovative Mobile Musician Machine Interfaces
- Music as an interaction element
- Mobile HCI and music composition
- Spontaneous music, mobility and audio sharing
- Mobile collaborative and distributed composition
- Mobile collaborative performance
- Mobile digital technology in music education
- Nomadic composition.

3 Aims

This workshop aims to promote dialogue on Mobile HCI and Sound, encourage research in the area and to share experiences to mutual advantage among people addressing the topic from different perspectives.

4 Intended Participants

We specifically wish to foster exchange and interaction among *participants from different communities,* including, but not limited to, designers, developers, computer scientists, musicians, evaluators, psychologists, and ethnographers, thus fostering cross-fertilization among research communities. Participants at all stages in their research are encouraged to participate. Research based on implemented systems will be particularly welcome.

5 Demonstrations

Demos and experiences based on implemented systems, (including those that integrate mobile technologies with fixed ones - e.g. large projection screens, synthesizers, input devices etc) are warmly welcome. It is anticipated that a surround sound system will be available if required for demonstrations. We particularly encourage demonstrations that allow individuals and small groups to gain hands-on experience.

Second International Workshop on Mobile and Ubiquitous Information Access

Fabio Crestani[1], Matt Jones[2], and Stefano Mizzaro[3]

[1] Dept. Computer and Information Sciences
University of Strathclyde
Glasgow, Scotland, UK
f.crestani@cis.strath.ac.uk

[2] Dept. Computer Science
University of Waikato
Hamilton, New Zealand
always@acm.org

[3] Dept. of Math and Computer Science
University of Udine
Udine, Italy
mizzaro@dimi.uniud.it

Abstract. The recent trend towards pervasive computing and information technology becoming omnipresent and entering all aspects of modern living, means that we are moving away from the traditional interaction paradigm between human and technology being that of the desktop computer. This shift towards ubiquitous computing is perhaps most evident in the increased sophistication and extended utility of mobile devices, such as mobile phones, PDAs, mobile communicators (telephone/PDA) and Tablet PCs. Advances in these mobile device technologies coupled with their much-improved functionality means that current mobile devices can be considered as multi-purpose information access tools capable of complex tasks. This Second Workshop on Mobile and Ubiquitous Information Access aims to be a forum for the presentation of current research and exchange of experiences into technological and usability aspects of mobile information access.

1 Motivations

The ongoing migration of computing and information access from the desktop and telephone to mobile computing devices such as PDAs, tablet PCs, and next generation (G3) phones poses critical challenges for research in information access and, in particular, for Information Retrieval (IR). These devices offer limited screen size and no keyboard or mouse, making complex graphical interfaces cumbersome. This change in information access devices is also reflected by a radical change in user groups and tasks. Most future users will have low levels of IT literacy and will not be information access professionals, but casual users. Therefore, these mobile devices will be used in situations involving different physical and social environments and tasks, and they will need to allow users to interact

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 529–530, 2004.

wherever he/she is and using whichever mode or combination of modes are most appropriate given the situation, their preferences and the task at hand. Furthermore, unlike traditional library or office settings, users of mobile information access devices will, typically, be subject to much higher levels of interruption and task switching, thus needing very different interface designs.

This workshop aims to be a forum for the presentation and discussion of current research and development of technological and usability aspects of mobile information access. The workshop is interesting for both researchers and practitioners of several different communities, such as information retrieval, digital libraries, HCI, mobile devices, and so on. A wide range of topics are relevant to the objectives and aims of the workshop, in particular:

- information retrieval and filtering;
- user modelling and personalisation;
- context awareness;
- new mobile devices;
- nomadic computing;
- ubiquitous computing;
- usability;
- ambient intelligence.

The workshop aims at addressing these topics both in terms of existing approaches and implementations, and in terms of theoretical foundations and emerging directions of research.

2 History

A workshop with the same name and theme was held last year at Mobile HCI 2003 in Udine, Italy. Two of the current workshop organisers were in the organising committee of that workshop. The workshop was very successful, both in terms of submissions (over 30 papers) and participants (over 40 participants). Selected papers were invited to submit a revised and extended version for inclusion in a volume published after the event [1].

3 Publication of Proceedings

The workshop proceedings are going to be made available online. In addition, while last year a volume was edited including selected papers of the proceedings, this year we envisage the preparation of a special issue of an internationally recognised journal. Selected papers accepted for presentation at the workshop will be invited to submit a revised and extended version.

References

1. F. Crestani, M. Dunlop, and S. Mizzaro, editors. *Mobile and Ubiquitous Information Access*, volume 2954 of *Lecture Notes in Computer Science*, Berlin, Germany, 2004. Springer-Verlag.

Mobile Communications Versus Pervasive Communications: The Role of Handhelds

Akseli Anttila[1], Amnon Ribak[2], Anxo Cereijo Roibás[3], Sabine Seymour[4], Sofia Svanteson[5], Scott Weiss[6], and David Williams[7]

[1] Senior Concept Designer, Nokia, Finland, akseli.anttila@nokia.com
[2] Researcher, Telecom & Media, IBM Research Lab in Haifa, Israel, ribak@il.ibm.com
[3] School of Computing, Mathematical & Information Sciences, Faculty of Management & Information Sciences, University of Brighton, a.c.roibas@brighton.ac.uk
[4] Founder and Chief Creative Officer, Moondial, sabine@moondial.com
[5] Interaction Designer, Ocean Observations, sofia@oceanobservations.com
[6] Principal, Usable Products Company, sweiss@usableproducts.com
[7] Independent UI Consultant, david.m.williams@libero.it

Abstract. The intent of this panel is to stimulate discussion around the design of HCI for handhelds and related applications in pervasive communication scenarios. The panel will create a provocative framework for future interactive communications and will generate a debate with the audience around the emerging and contentious topic of using mobile devices as main interfaces in a ubiquitous communication experience. It is expected that the panelists will promote inter-disciplinary and multidisciplinary discussion about a range of HCI issues including: management of multiple user interactions, interaction between mobile phones and other devices, the development of usable design models for convergent media, special HCI considerations around media sharing, new applications and new forms of interactive content more suitable in pervasive communication contexts. Other ethical, legal, business and social implications of ubiquitous communications may also be raised. Short position statements will be followed by moderated debate between panelists that will be open to questions from the floor stimulating high interaction with the audience. The panel will conclude with a summarization of the most relevant outcomes of the discussion.

1 Introduction

Incoming pervasive communication scenarios anticipate how mobile phones, as personal interfaces interconnected with other surrounding platforms (e.g., iTV, PC's, PDA's, in-car-navigators, smart-house appliances, etc.), will strongly contribute to create this communication ubiquity.

These new scenarios will imply the need to rethink new kinds of services and applications and of course new forms of content. For example, handsets are becoming tools for creation, editing, and diffusion of personalized and personal multimedia

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 531–535, 2004.

content on a ubiquitous network. It is not difficult to imagine how these mobile devices will contribute to the scenario of TV everywhere, or better, iTV.

Some of the issues that need to be addressed during the proposed debate are: what are the immediate and long term advantages for mobile users in the future? Will handsets contribute to improve accessibility to pervasive systems, or on the contrary, will ubiquitous contexts make the interaction formats less usable? What are the core issues regarding usability and accessibility for input-output devices to ubiquitous systems? How will elderly people, young children and the physically disabled respond to a ubiquitous model based on interaction through the mobile phone? What are the interoperability issues that need to be addressed? How can ubiquitous systems gain from the application of context awareness to mobile services? What other paradigms exist beyond having contextualized access to information? Will customization become a must? What are the challenges for handheld manufacturers, content producers, broadcasters and operators in this new scenario? Most SMS TV functions are for chat purposes only or peer-to-many communication, so what are new scenarios for SMS applications regarding iTV? What new mobile applications and technology could be developed regarding interaction with iTV? Will a new mobile users' community be created? Will this community communicate and exchange content one-to-many? Would content be beyond SMS toward a more rich media?

2 Significance and Timeliness

This panel explores new scenarios in ubiquitous communications, trying to shed light into new related experience models and new forms of content. Convergent media are gaining more prominence especially regarding new forms of content for mobile devices. Each interface (PC, iTV, mobile phone, PDA, car navigator, etc.) has its own characteristics both from the interactive and the technical points of view. HCI designers therefore need to know the most suitable service formats and the distinctive interaction patterns for each interface enhancing the interoperability of all its features and optimizing its usability. Technicians must be aware of the latest technologies that could be required in these ubiquitous systems, content providers have to identify new forms of content that will be more suitable in the context of convergent media, network operators need to spot new applications and services that will satisfy better users' needs and expectations in these new environments.

3 Panelists' Position Statements

3.1 Akseli Anttila

Akseli Anttila is a senior concept designer at Nokia Research Center, Finland. His current work focuses on the convergence of traditional media with mobile and connected devices, and he has designed and studied applications for the cross-media use of TV and radio with the mobile handset. He is also a Ph.D. candidate at the Univer-

sity of Art and Design, Helsinki on "Machine mediated communication and music enjoyment."

Position: The mobile and media industries are faced with the challenge of creating new applications and devices for the enjoyment of media in a mobile context. Such yet to be designed applications must take into account the realities of the mobile context of use. Changes in the users' physical and social surroundings in enjoying media in mobile contexts impose various new constraints on the affordances of mobile media applications and devices. Users' ability to focus, interact, and manage interruptions in the enjoyment of media need to be considered when creating usable media for mobile contexts. Furthermore, applications must take into account the need for user creation and modification of content. Particularly, user interface design of future mobile media applications should place emphasis on users' actions within their social network, and allow users to repurpose basic device functionality to their own (social) enjoyment, since one of the primary purposes of the mobile convergent device will be communication with others.

3.2 Ammon Ribak

Amnon Ribak joined IBM Research in 1990 as a Computational Linguist. His work focuses on "pleasant interfaces", promoting the use of speech, multimedia, Instant Messaging and other lightweight user interfaces to people, knowledge and applications in the enterprise.

Position: Future devices will be mostly interfaced with voice, but new forms of typing and writing will emerge as well. While devices get smaller and smaller, they will use their immediate environments for I/O purposes. Here are some examples:

Speech Interface: BlueTooth microphones, attached to the shirt, will form a microphone array, assisting in beam-forming and noise reduction. Distributed Speech Recognition, extracts speech-recognition features and passes them to the recognition engine over an IP channel. Cameras will use lip-reading to improve speech-recognition accuracy.

GUI's: devices will use new means of projecting their graphical displays on nearby walls or paper, as well as on glasses, car-windows, or TVs, forming see-through interfaces that merge into the environment.

Data Entry: Virtual keyboards and wireless external writing-pads will recognize handwriting.

While the environment will serve as an extension to the handheld device for interface purposes, it will also provide context to the device: buildings will know who is in them, sensors will monitor body conditions, cars and trucks will advise on road conditions and driver's fatigue, etc.

3.3 Anxo Cereijo Roibás, Ph.D.

Anxo Cereijo Roibás is a Senior Lecturer at the School of CIMS of the University of Brighton, Contract Professor at the Politecnico di Milano University and at the Uni-

versity of Milan. He also collaborates with Vodafone and Nokia in the design of HCI for contexts of ubiquitous computing in multi-access & multi-channel scenarios of convergent media.

Position: The incoming Ubicomp communication scenarios involving cross-platform customer technologies (ranging from iTV, radio, music, and mobile phones to portable or wearable information devices) requires an original way of thinking about the interactive user experience, which indicates the need to design novel ubiquitous and mobile services and products that will address the new demands, needs, and potentials of mobile users. At the same time, it becomes crucial to develop new experience models and their social, cultural and regulatory implications, exploring new and relevant interactive models and paradigms, rethinking the added-value of interactivity for the users across immersive and multi-user environments and context awareness applications and consequently creating and refining new forms of rich smart cross-media content (including virtual objects, multi-user environments and immersive, intelligent content for haptic and sensor-based interfaces and animated content). I envisage original content that flows from multiple sources and over different pipelines, which is processed, purposed and enriched for different contexts and different audiences, which is displayable on a wide range of devices, which is multi-channel, and multi-format, multi-accessible, flexible and cost-effective.

3.4 Sabine Seymour

Sabine Seymour founded and is the Chief Creative Officer of Moondial, started in New York in 1998, and in Vienna, Austria since 2004. Moondial, Fashionable Technology Collective, focuses on the convergence of fashion, wearable & wireless technologies, design, and architecture in the areas of extreme sport and fashion/style. Sabine is currently a member of the ISEA 2004 IPC for the Wearable Experience Section. She is an Adjunct Professor and Design Fellow at Parsons School of Design.

Position: Technology has enabled a greater degree of personalization in fashion. Marrying the elements of extreme sports and urban couture with smart and high performance textiles, our collaborative has developed a functional and fashionable athletic jacket concept and prototype. Digital pictures from a camera phone or PDA can be transmitted directly, or downloaded wirelessly from the internet, to a display embedded in the jacket.

3.5 Sofia Svanteson

Sofia Svanteson is an Interaction Designer, CEO and founder of Ocean Observations. She studied Media Production and Human Computer Interaction at the Royal Institute of Technology in Stockholm, where she wrote one of the world's first papers on usable mobile GUI's. With a passion for usability, usefulness and beauty, Sofia has, together with her colleagues, raised Ocean into a mobile expertise company with clients such as Samsung, 3, Motorola and Orange in its portfolio.

Position: What happens in Tokyo today is often the future of the rest of the world, especially in terms of mobile handsets and their applications and services. At this very moment, people of a wide range of ages are running (or driving) around Tokyo using their mobile phones participating in the location based, community driven, and non-fighting mobile game Mogi Mogi (`http://mogimogi.com/`). The game is about collecting all items within a certain series, placed all around Tokyo, using a GPS enabled mobile phone. Trading items (duplicates) with other players is a must in order to complete collections and win the game. This application combines a mobile UI, a web UI, and an instant messaging component.

3.6 Scott Weiss, Chairperson

Scott Weiss is the Principal of Usable Products Company, an ease-of-use consultancy focused on mobile device usability and design. Usable Products' clients include Sprint PCS, LG Electronics, Microsoft, and Intel. Scott's work experience includes career positions at Apple, Microsoft, Sybase, and Autodesk. He is the author of "Handheld Usability" (Wiley: 2002).

Position: Converging technologies spark the fancy of many, but delight few due to poor user interface design and usability. As the camera and the mobile telephone have converged, usability has deteriorated. Millions are taking low quality, out-of-focus pictures—and they cannot figure out how to send them to each other. As the MP3 player and the mobile telephone converge, users cannot figure out how to find their favorite albums, instead giving up and purchasing the superior iPod MP3 player. Convergence is leading to divergence. My role in this panel will be to ask the panelists to not only share their ideas for the future, but to rally them and the audience around excellent design processes, so that the realization of these designs will be as wonderful as their creativity.

3.7 David Williams, Ph.D.

David is founder of a small mobile user interface design consultancy based in Milan, Italy. Current projects include a co-design leadership role with Motorola and its key customers. David has a Ph.D. in Multimedia User Interface design and 8 years experience in R&D, Development and Design Consultancy roles.

Position: Ubiquity is not about technology but about people being empowered to do new (and old) things in new places, moving and stationary. "Everyday" mobile usage is not about complex interactions or experiences; rather it is fragmented, capricious and often interrupted: an interaction style which reflects the environments that people are living in. In order to support ubiquity, mobile user interfaces must appeal as much to the emotional as the logical. The interface and information must be dynamically matched to the users' situation and needs in a way which is both engaging and useful. The interface becomes mobile – moving from phone, to wall, to kiosk, to PDA. How should the HCI community face ubiquity? First, emotive and intimate design. Second, contextual interaction. Third, designing together.

The Myth of the ‘Martini Solution’

Richard Harper[1], Abigail Sellen[1], Tim Kindberg[2], Phil Gosett[3], and Kaisa Väänänen-Vainio-Mattila[4]

[1] MSR Cambridge, UK
[2] Hewlett-Packard Labs, Bristol, UK
[3] Vodafone R&D, UK
[4] Tampere University of Technology and Nokia Corporation, Finland

1 Panel Focus

Some ten or fifteen years ago there was a well-known TV advert shown in the UK that implied that cool people drank Martini ‘anytime, anyplace’. Since then, in the UK at least, the term ‘Martini solutions’ has been used to describe the products of the mobile industry, where the claim is made that these products are usable anytime, anyplace. Further, that the key value of mobile products and services is that they allow people to do things in time and space that they could never do before.

This proposes a panel that will explore the tension between designing for ‘anytime-anywhere’ and designing for 'situated use'. It will consider the proposition that too little attention has been given to situated use both by the mobile industry and in academic mobile HCI research resulting in applications that are neither good 'anytime, anywhere' nor good for particular places.

The panellists will present their own views on the conceptual distinction between situated and anytime, anyplace solutions, and will bring to bear their own very distinct organisational and scientific backgrounds on product technology, design and usability methodologies.

1.1 Timeliness

This issue is particularly salient at the current time as the mobile industry tries to move beyond telephony and messaging towards what it calls ‘3g offerings’.

2 Biographical Sketches

Richard Harper has been at the forefront of academic sociological research into the role and shaping of mobile technologies, and is editor of *The Wireless World* (Springer, 2001) and *Inside Text: Cultural and design perspectives on SMS* (Kluwer, forthcoming). He now jointly leads Microsoft’s Socio-Digital Systems group at Cambridge.

S. Brewster and M. Dunlop (Eds.): MobileHCI 2004, LNCS 3160, pp. 536–537, 2004.

Tim Kindberg is senior research scientist at HP Labs Bristol and was formerly a senior lecturer in Computer Science at Queen Mary, University of London. He has many interests in the areas of mobile and ubiquitous computing including the link between the physical and the digital world, and trust and security in mobile interaction. He is co-author of the textbook "Distributed Systems: Concepts and Design."

Abigail Sellen is a cognitive psychologist at Microsoft Research in Cambridge who studies people with an eye toward the design of new technology. Her interests are diverse, including: reading, paper use, Web use, videoconferencing, human memory, computer input, and information capture. Her latest work has implications for the design of new mobile devices and applications most particularly in the area of family life.

Phil Gosett leads work on consumer insight and the social aspects of the Vodafone's future vision within its R&D function in Newbury, England. His interests range from building a fraud detection engine for mobile billing through to explorations of the cultural differences that affect the use of mobile devices around Europe.

Kaisa Väänänen-Vainio-Mattila manages to combine a professorship of usability and user-centered design at Tampere University of Technology with being the director of Consumer Insights at Nokia Corporation. Much of her research has focused on the topic of situated (or contextual) use in the design of mobile devices, applications and services, and in the development of pragmatic design processes for both 'anytime, anywhere' and situated solutions.

Author Index